

# **Dissertation**

submitted to the  
Combined Faculties for the Natural Sciences and for Mathematics  
of the Ruperto-Carola University of Heidelberg, Germany  
for the degree of  
Doctor of Natural Sciences

presented by

Diplom-Biologe Thomas Conrad  
born in: Neustadt an der Weinstrasse  
Oral-examination: 28.06.2011

**Chromatin regulation by the histone acetyltransferase MOF in**  
***Drosophila***

Referees:

Dr. Francois Spitz  
Prof. Dr. Gabriele Petersen



For Katrin

# Table of Contents

<b>Acknowledgements</b>	<b>6</b>
<b>Zusammenfassung</b>	<b>7</b>
<b>Summary</b>	<b>9</b>
<b>1. Introduction</b>	<b>11</b>
<i>Preface</i>	12
<i>1.1 Chromatin</i>	12
1.1.1 The nucleosome	12
1.1.2 Higher order chromatin structure	13
1.1.3 Remodeling chromatin	15
1.1.4 Nucleosome assembly / disassembly	15
1.1.5 ATP dependent chromatin remodeling	17
1.1.6 Histone variants	20
1.1.7 Posttranslational modification of histones	22
1.1.8 Noncoding RNAs and chromatin	26
1.1.9 Functional states of chromatin	29
<i>1.2 Dosage compensation in Drosophila</i>	30
1.2.1 Sex chromosomes and the need for dosage compensation	30
1.2.2 The MSL complex	31
1.2.3 Programming of the male X chromosome	32
1.2.4 Transcriptional activation	34
1.2.5 Other factors	35
1.2.6 Integration of MOF into the dosage compensation complex	36
<b>2. Aims and Objectives</b>	<b>39</b>
<b>3. Results</b>	<b>43</b>
<i>3.1 MOF is responsible for H4K16ac on all chromosomes</i>	44
<i>3.2 MOF is essential for female survival</i>	45
<i>3.3 Most active genes are targeted by MOF and H4K16ac in both sexes</i>	46
<i>3.4 The MOF N-terminus controls DCC function</i>	52
<i>3.5 The N-terminus constrains MOFs enzymatic activity</i>	60
<i>3.6 The Zn finger of MOF is required for targeting to the X and autosomes</i>	61
<i>3.7 The chromobarrel domain is required for global H4K16ac</i>	63
<i>3.8 The chromobarrel domain triggers acetylation after nucleosome binding</i>	70
<b>4. Discussion</b>	<b>73</b>
<i>4.1 MOF is the major H4K16 specific HAT in Drosophila</i>	74
<i>4.2 MOF controls dosage compensation via its N-terminal domain</i>	75
<i>4.3 The MOF chromobarrel domain controls H4K16ac genomewide</i>	77
<i>4.4 Summary</i>	80

<b>5. Ongoing Projects</b>	<b>81</b>
5.1 <i>The mechanism of dosage compensation</i>	82
<b>6. Published Work</b>	<b>87</b>
<b>7. Materials and Methods</b>	<b>89</b>
7.1 <i>Chromatin immunoprecipitation (ChIP)</i>	90
7.1.1 ChIP from salivary glands	90
7.1.2 ChIP from whole larva	90
7.2 <i>RNA isolation and quantitative PCR analysis</i>	91
7.3 <i>Fly stocks and crosses</i>	92
7.4 <i>Immunostaining of polytene chromosomes and confocal microscopy</i>	92
7.5 <i>Generation of protein extracts and western blotting</i>	93
7.6 <i>Expression and purification of Drosophila recombinant proteins with the Baculovirus system</i>	94
7.7 <i>HAT-Assay on nucleosomal template</i>	94
7.8 <i>Filter binding HAT-assay</i>	95
7.9 <i>Electrophoretic mobility shift assay (EMSA)</i>	95
7.10 <i>Immunoprecipitation of MOF constructs from SL-2 stable cell lines</i>	95
7.11 <i>ChIP-seq data analysis</i>	96
7.11.1 Data processing	96
7.11.2 Detection of MOF-bound and H4K16-acetylated genes	97
7.11.3 Gene expression profiling	97
<b>8. References</b>	<b>99</b>
<b>Appendix A</b>	<i>“Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in Drosophila melanogaster.” Mol Cell Biol 28, 3401-3409.</i>
<b>Appendix B</b>	<i>“The nonspecific lethal complex is a transcriptional regulator in Drosophila.” Mol Cell 38, 827-841.</i>
<b>Appendix C</b>	<i>“The MOF chromobarrel domain controls genomewide H4K16 acetylation and spreading of the MSL complex” In review in Genes and Development.</i>

## **Acknowledgements**

The time of my PhD was certainly one of the most valuable experiences in my life, and I want to first of all thank Asifa for giving me the chance to be part of her endeavour over these last five years. I also want to thank the members of my TAC Jürg Müller, Gabriele Petersen and Francois Spitz for their guidance and advice during the TAC meetings.

I am especially grateful to Jop, who inherited me a great project and introduced me into the fly world, and Steve and Sascha for the great time we had in the Bars of Heidelberg in these early years. And of course I want to thank Sunil for all the discussions, for the work that we did together and for the great clubbing afterwards. Then Erinc and Ibrahim joined the lab, bringing in some exciting discussions and a special sense of humor that helped to cope with the tough times of the PhD experience. Of course I also want to thank Ghislaine for the energy that she brings into this lab. Certainly, I could not have done this work without Herbert, who will never give up, no matter what it takes to get a HAT assay to work! The same goes for Florence and Juanma, who have made an incredible effort to get all this data analyzed. And I want to thank Jacky and Johanna for their friendship, it is so good to know that one is not alone in this.

Finally I want to apologize to my family and friends for the little time that I could spend with them, and tell them how thankful I am for their love and support. During these five years I was absorbed in the surreal world of this lab space, and most of all I am so grateful that Katrin has always been with me, even from the distance. After all, this PhD has also shown me what the most important things in my life are.

## Zusammenfassung

Geschlechtsbestimmung in *Drosophila* geht, wie auch in Säugetieren, einher mit der ungleichen Verteilung von Geschlechtschromosomen. Männliche Fruchtfliegen tragen ein X und ein Y Chromosom, im Gegensatz zu zwei X Chromosomen in Weibchen. Um die negativen Konsequenzen chromosomaler Aneuploidie zu verhindern haben Fruchtfliegen im Verlauf der Evolution ein Dosiskompensationssystem entwickelt, welches die Transkription vom männlichen X Chromosom zweifach erhöht und somit das Transkriptionsniveau der zwei weiblichen X Chromosomen erreicht. Diese Aktivierung wird durch den „dosis compensation complex“ (DCC) bewirkt, ein Ribonucleoproteinkomplex bestehend aus fünf „male specific lethal“ (MSL) Proteinen und zwei nicht kodierenden „RNAs on the X“ (roX). Der DCC bindet an hunderte von Zielorten auf dem männlichen X Chromosom, wo er die Hyperazetylierung von X chromosomalem Chromatin an Lysin 16 von Histon H4 (H4K16ac) vermittelt. Diese Histonmodifikation ist mit einer offenen, permissiven Chromatinstruktur assoziiert, und es wird angenommen dass ihre Anreicherung auf dem männlichen X Chromosom für die zweifache Aktivierung der X chromosomalen Transkription während der Dosiskompensation notwendig ist. Ungeachtet dessen ist der exakte Mechanismus durch den X chromosomale Transkription in Männchen aktiviert wird unbekannt. Verantwortlich für die Hyperazetylierung des männlichen X Chromosoms ist die Histonazetyltransferase „males absent on the first“ (MOF), welche Teil des DCC ist. Es wurde kürzlich gezeigt dass MOF eine zusätzliche Funktion in autosomaler Genregulation hat, da es an tausenden von Genpromotoren als Teil des „non specific lethal“ (NSL) Komplex gefunden wurde. Zu welchem Grad H4K16ac an autosomalen Genen allerdings von MOF abhängig ist, und wie die Verteilung von MOF zwischen den beiden Komplexen reguliert wird ist nicht bekannt.

Im Verlauf meiner Promotion habe ich genetische, biochemische und genomweite Verfahren verwendet um einen weiten Bereich von Fragen bezüglich der Funktion von MOF in autosomaler Genregulation und Dosiskompensation, der Bindung des DCC an X chromosomale Gene, und des Mechanismus der transkriptionellen Aktivierung X chromosomaler Gene während der Dosiskompensation zu beantworten. Neben Beiträgen zu anderen Arbeiten, darunter Untersuchungen zur Rolle der H3K36 spezifischen Methyltransferase HypB/Set2 in der Dosiskompensation, sowie zur Rolle von MOF für die Funktion von NSLs, habe ich diese Fragen im Kontext zweier Projekte adressiert.

Im Verlauf dieser Arbeiten konnte ich zunächst zeigen dass MOF in männlichen und weiblichen Fliegen genomweit für H4K16ac verantwortlich, und ein essentielles Gen in

Weibchen ist. Ich zeige dass der *Drosophila* spezifische unstrukturierte N-terminus von MOF für den Aufbau des DCC auf dem männlichen X Chromosom erforderlich ist und gleichzeitig die enzymatische Aktivität von MOF hemmt. Der N-terminus kontrolliert daher MOFs Funktion für die Kompensation des X Chromosoms. Weiterhin habe ich die biologische Funktion der von Hefe bis Mensch konservierten Chromobarrel Domäne im MOF protein aufgezeigt. Überraschenderweise führt die Mutation der Chromobarrel Domäne, welche für die Interaktion von MOF mit roX RNA nötig ist, zu einer dramatischen Abnahme von H4K16ac auf allen Chromosomen. Ich konnte somit zum ersten Mal eine biologische Funktion der MOF Chromobarrel Domäne *in vivo* zeigen, welche darin besteht H4K16ac nach der Rekrutierung von MOF zum Chromatin auszulösen.

In einem parallelen Projekt, welches darauf abzielt den Mechanismus der Dosiskompensation zu erhellen, wollte ich bestimmen an welchem Schritt des RNA Polymerase II Transkriptionszyklus Dosiskompensation in Fliegen agiert. Hiefür habe ich genomweite Bindungsprofile von RNA Polymerase II in Speicheldrüsen von männlichen und weiblichen Larven, und von Larven mit beeinträchtigter Dosiskompensation hergestellt. Diese Analyse zeigt dass RNA Polymerase II auf dem männlichen X Chromosom ungefähr zweifach angereichert ist, einschliesslich an Genpromotoren. Dieses Ergebnis legt nahe dass Dosiskompensation auf der Ebene der Transkriptionsinitiation arbeitet, was einen bedeutenden Fortschritt für unser Verständniss der Dosiskompensation darstellt.

## Summary

Like in mammals, sex determination in *Drosophila melanogaster* involves an unequal distribution of sex chromosomes, with male flies carrying an X and a Y chromosome, as compared to two Xs in females. To prevent the deleterious effects of chromosomal aneuploidy, flies have evolved a dosage compensation system, which upregulates transcription from the single male X chromosome to match transcript levels produced from the two female Xs. This transcriptional activation is achieved by the dosage compensation complex (DCC), a ribonucleoprotein complex consisting of five male specific lethal proteins (MSL) and two non coding RNAs on the X (roX). The DCC is physically tethered to hundreds of target loci along the male X chromosome, where it promotes hyper-acetylation of X-linked chromatin at Lysine 16 of histone H4 (H4K16ac). This histone mark is associated with an open, permissive chromatin structure, and its enrichment on the male X chromosome is thought to be required for the twofold increase in X-linked transcription during dosage compensation. However, the exact mechanism by which X-linked transcription is activated in males is still unknown. Responsible for hyper-acetylation of the male X chromosome is the histone acetyltransferase males absent on the first (MOF), which is part of the DCC. Recent studies have shown that MOF plays an additional role in autosomal gene regulation, as it has been found at thousands of autosomal gene promoters as part of the non specific lethal (NSL) complex. However, to what extent H4K16ac at autosomal genes is MOF-dependent, and how MOF is differentially distributed between the two complexes is currently unknown.

During the course of my PhD, I used genetic, biochemical, and genomewide approaches to address a wide range of questions, concerning MOF functions in autosomal gene regulation and dosage compensation; the DCC recruitment process to X-linked target genes; and the mechanism of transcriptional upregulation of X-linked genes during dosage compensation. Besides other contributions, investigating the role of the H3K36 specific methyltransferase HypB/Set2 during MSL targeting and dosage compensation, as well as the role of MOF for NSL function at autosomal promoters, I was addressing these questions in the context of two main projects.

During the first one of these, I have been able to show that MOF is responsible for genomewide H4K16ac in male and female flies, and that MOF is an essential gene in females. I demonstrated that the *Drosophila* specific unstructured N-terminus of the MOF protein is required for assembly of the DCC on the male X chromosome, and at the same time constrains MOFs HAT activity. The N-terminus therefore controls MOFs function in X

chromosome compensation. I was furthermore able to reveal the biological role of the chromobarrel domain, which is conserved from yeast to human. Unexpectedly, disruption of the MOF chromobarrel domain, which has been shown previously to be required for MOF interaction with roX RNAs, led to a dramatic loss of H4K16ac from all chromosomes. Accordingly, I showed that the chromobarrel domain serves to trigger H4K16ac after the recruitment of MOF to its chromatin targets, revealing for the first time a biological role of this domain *in vivo*.

In a parallel project, to work towards unraveling of the dosage compensation mechanism, I wanted to identify the step in the RNA PolII transcription cycle at which dosage compensation operates in flies. To this end, I generated genomewide profiles of RNA PolII in 3<sup>rd</sup> instar larva salivary glands from male and female flies, and from male flies with disrupted dosage compensation. Strikingly, we find that the density of PolII is approximately twofold elevated on the male X chromosome as compared to autosomes, including X-linked promoters. This data suggests that dosage compensation operates via enhanced transcription initiation, which constitutes a major advance in our understanding of the dosage compensation process.



# **Introduction**

## Preface

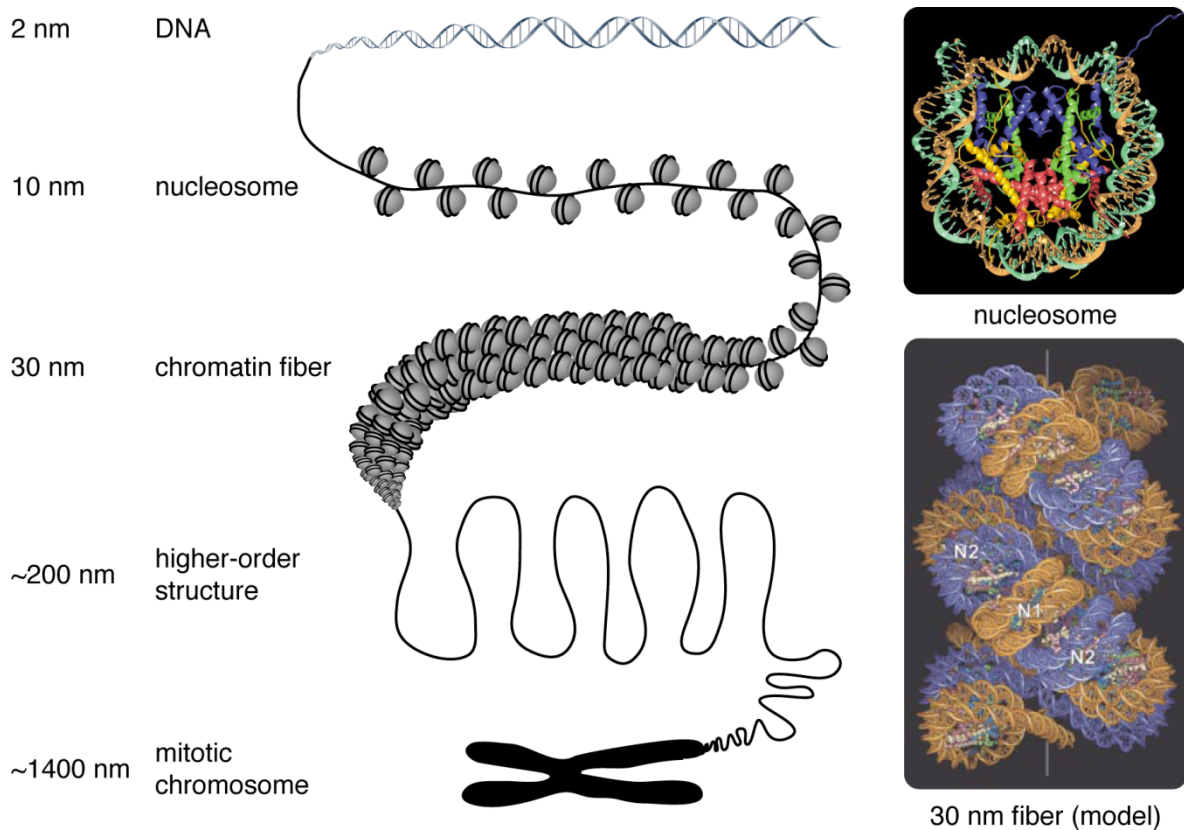
In metazoans, the same genetic information stored inside the nucleus of each somatic cell can give rise to highly specialized cell types, specific for certain tissues and/or developmental stages. These functional states are the result of distinct patterns of gene expression. Accordingly, at any given time point during its life cycle, each cell only realizes a subset of the transcriptional and developmental potential encoded in the genome as a whole. A fundamental requirement for the formation of a multicellular organism is therefore the ability to read the genetic code in a context dependent manner, and to faithfully maintain transcriptional states from one cell generation to the next. At the basis of this “epigenetic” memory lies the compaction of the genomic sequence into “chromatin”. This nucleoprotein complex allows for the activation or inactivation of any genomic locus through an alteration in chromatin structure, mediated by a complex interplay of transcription factors, noncoding RNAs, and histone modifications. This introduction will give a brief overview over the principles that underlie this sophisticated process.

## 1.1 Chromatin

### 1.1.1 The nucleosome

The fundamental unit of chromatin is the nucleosome with 147bp of DNA wrapped around an octamer of histone proteins in  $1\frac{3}{4}$  left-handed superhelical turns [1], leading to a 5- to 10-fold compaction of the DNA. The center of the octamer consists of two dimers of histone H3 and H4 while a dimer of H2A and H2B lies on either side of the nucleosome, as depicted in Figure 1 [2]. Common to all four histones is the central histone fold domain (HFD), consisting of three  $\alpha$ -helices ( $\alpha 1$ ,  $\alpha 2$  and  $\alpha 3$ ) separated by two loops (L1 and L2), which is conserved throughout eukaryotic and archeal lineages (for a review see [3]). The core histone octamer is only stable in the presence of DNA or at very high salt concentrations, due to the highly basic charge of all four core histones. Core nucleosomes are arranged on the DNA like “beads on a string” with 10-80pb of linker DNA between them [4], and one copy of the linker histone H1 can bind to and protect 20bp of this linker DNA against micrococcal nuclease digestion [5]. Histone H1 is not required for the assembly of the core nucleosome itself but is thought to facilitate formation of the next higher level of chromatin structure, the 30nm fiber [6, 7]. In vivo, depletion of H1 leads to reduced nucleosome repeat length and defects in the compaction of mitotic chromosomes, and severe depletion of H1 is fatal in mice

and *Drosophila* [8, 9]. The unstructured tails of the four core histones are not required for the formation of the mononucleosome but do affect inter-nucleosomal interactions and the folding of the 30nm fiber, as well as the recruitment of chromatin modifying factors and the transcriptional machinery. All of these functions are extensively modulated by covalent posttranslational modifications of the histone tails, as will be discussed in more detail below.



**Figure 1. Organization of DNA in the eukaryotic nucleus.**

The crystal structure of the mononucleosomes with histone H2A, H2B, H3 and H4, adapted from [10], and the suggested 30 nm fiber structure from [11].

### 1.1.2 Higher order chromatin structure

Higher order chromatin structure may be defined as any group of nucleosomes that adopt a reproducible conformation in 3D [12]. In analogy to protein structures, the chain of nucleosomes might form secondary and tertiary structures mediated by inter-nucleosomal interactions. Although the primary structure of chromatin is a simple repetition of nucleosomes, the situation is complicated by the existence of histone variants, by post-

translational modifications of the unstructured histone tails and by variable spacing between nucleosomes, which all might impact on chromatin secondary structures. To directly prove the existence of such structures has so far proven difficult, and recent attempts using cryo-electron microscopy have failed to support the existence of a 30nm fiber *in vivo* [13]. However, extensive work over the last decades on isolated and reconstituted chromatin has revealed some reoccurring patterns of secondary structure formation. Free DNA is highly negatively charged and self-repulsive, and this charge is not fully compensated by the basic histone proteins. Chromatin structure *in vitro* is therefore highly responsive to changes in the ionic environment [14]. With increasing ionic strength, the beads on a string conformation visible in electron micrographs is getting progressively compacted into a fiber with ~30nm in diameter (Figure 1). Two alternative nucleosomal arrangements have been proposed for the 30nm fiber. In the solenoid one-step helix model, the linker histone H1 is placed inside the helix and the linker DNA connects two consecutive nucleosomes within the same stack. Per turn, 6-8 nucleosomes coil around a central cavity constituting the superhelix [15]. In contrast, in the two-start model the linker DNA is found on the outside of the superhelix, connecting nucleosomes of consecutive stacks [16]. The two-start model has been supported by a study that used the interaction between an acidic patch within the flat face surface of the H2A-H2B dimer and the N-terminal tail of histone H4 for di-sulfite crosslinking of corresponding cysteine replacement mutants [10, 17]. The crosslinked fibers were prone to digestion by restriction enzymes specific for the linker DNA sequences, suggesting an orientation towards the outside of the fiber. In further support for a two-start helix, the recent successful crystallization of a tetranucleosome revealed a complex that assumed a zigzag arrangement with interactions between nucleosomes 1 and 3 and between 2 and 4, with the linker extending between them [11]. Finally, a new technique termed EMANIC (EM-assisted nucleosome interaction capture), in which formaldehyde crosslinked chromatin fibers are dispersed in low salt and analyzed by electron microscopy, has provided evidence for a mixed fiber architecture predominated by the zigzag configuration and interrupted by senoidal segments [18]. Interestingly, computer modeling has shown that mixed conformations allow for a higher degree of compaction, and better match the calculated nucleosome packaging *in vivo*. The existence and form of hierarchical structures above the 30nm fiber are a matter of intensive study and, due to the ~250nm resolution limit of light microscopy, remain largely elusive to date. Muller et al. reported that ~400kb of non translated chromatin occupy roughly spherical volumes with about 250nm in diameter [19]. The corresponding ~2000 nucleosomes, if present as 30nm fiber, would extend to a length of 2μm, suggesting extremely

tight packaging of untranscribed chromatin. It will be interesting to see if the recent development of subdiffraction light microscopy will reveal any internal structures in these compact regions.

### **1.1.3 Remodeling chromatin**

During development and differentiation cells show remarkable transcriptional plasticity. Accordingly, since the interpretation of all genetic information relies on direct binding of regulators and the transcriptional machinery to DNA, the packing and unpacking of chromatin is a tightly controlled process, specific for loci harboring genes that need to be activated or repressed at any given time point as a response to environmental cues. Access to the DNA is not only modulated by packaging into higher order structures like the 30nm fiber, but also by the positioning of nucleosomes on the DNA. Nucleosomes occlude transcription factor binding sites and hinder RNA PolII while it is travelling through the gene [20], and *in vitro* transcription of a chromatinized template is severely slowed down, as compared to naked DNA [21]. However, while promoter regions tend to be depleted of nucleosomes *in vivo* in order to allow transcription factor and Pol II binding to DNA, most genes carry nucleosomes in their transcribed regions [22, 23]. The active regulation of chromatin structure and the ability of RNA PolII to overcome the obstacles imposed by the chromatin template are the result of a complex interplay between nucleosome assembly and disassembly by histone chaperones, ATP-dependent remodeling, replication independent incorporation of histone variants and post transcriptional modifications of the histones. Each of these mechanisms will be discussed below. An additional modification that impacts on chromatin structure and transcription is DNA methylation; however, since this modification plays a minor role in *Drosophila* it is beyond the scope of this brief introduction to chromatin biology.

### **1.1.4 Nucleosome assembly / disassembly**

In order to prevent aggregation and potentially deleterious effects of their highly positive charge, histones, when not in association with DNA, are under most circumstances bound to histone chaperones. These proteins not only prevent unwanted interactions with other factors, as previously believed. Histone chaperones can act alone or as part of larger protein complexes to link the incorporation of histones into DNA and the removal of histones

off DNA to processes like replication, transcription and DNA repair [24, 25]. Many histone chaperones have higher affinities for certain histones and are specific for distinct processes, and they are often able to not only catalyze the assembly of histones into or disassembly from DNA but also the reverse reaction. Some histone chaperones can on their own bind and transfer histones, such as Asf1, while others are part of multichaperone complexes with more than one chaperone subunit as in the case of the CAF-1 complex. Initially, nucleosomes are established during S-phase in a step-wise process in which the histone chaperones CAF1 and Asf1 deposit H3/H4 tetramers onto the DNA. It is currently believed that Asf1 acts as a histone donor for CAF-1 during this process, whereas another histone chaperone, Nap1, later adds H2A/H2B to build the octasomes [26, 27]. Structural analysis has revealed that Asf1 binds to the same interface in dimeric H3/H4 that also mediates interaction between two H3/H4 dimers, and it has been demonstrated *in vitro* that Asf1 is able to disrupt the H3/H4 tetramer [28]. A role for Asf1 in nucleosome disassembly has also been supported by studies in *S. cerevisiae*, showing that Asf1 is required for efficient nucleosome eviction during transcription [29]. However, 95% of the H2A/H2B dimers but only 5% of H3/H4 dimers are exchanged during transcription, suggesting that the Pol II complex is able to transcribe through the H3/H4 tetramer [30]. Transcription coupled disassembly of H2A/H2B is mediated by FACT (facilitates chromatin transcription), a multichaperone complex consisting of Spt16 and SSRP1, which also promotes reassembly behind the elongating Pol II. This reassembly is crucial as exemplified by deletion of *spt6* in yeast cells. This histone chaperone also promotes nucleosome reassembly after Pol II passage, and loss of *spt6* leads to aberrant transcription initiating from cryptic promoters in the transcribed regions [31]. Many histone chaperones also provide histone binding capacity in the context of large enzymatic complexes, such as actin-related protein (Arp) 7 and 9 in the SWI/SNF chromatin remodeling complex, or Arp4 in the INO80 complex, where they can be involved in transcription regulation, DNA repair and other processes. In addition, histone chaperones can also link histone modifying activities to their substrates, as in the case of Vps75 and the H3K56 specific HAT Rtt109 in *S. cerevisiae* [32], or Arp4, which recruits the NuA4 complex to  $\gamma$ H2AX in budding yeast [33]. Many other histone chaperones are involved in the propagation of epigenetic marks or in processes such as transcriptional memory, and the physical and functional network of histone chaperone interactions is under active investigation. Since the aim of this introduction is to provide a brief overview over the basic principles of chromatin biology, I recommend recent reviews for further reading on this fascinating field of research [24, 25].

### 1.1.5 ATP dependent chromatin remodeling

The exact positioning of nucleosomes on the DNA can have a profound impact on gene expression, since it can dramatically alter the accessibility of the DNA for transcription factors. Although preferential positioning of nucleosomes over certain DNA sequences has been clearly demonstrated *in vitro*, there is still an ongoing debate as to whether nucleosome positions *in vivo* are encoded in the genome or the result of sequence independent remodeling [34-36]. In any case, the ability to actively move nucleosomes over favored or unfavored positions on the DNA sequence is required for basically all chromatin related processes, like DNA replication, recombination, repair and transcription [37]. ATP dependent chromatin remodeling enzymes are highly conserved proteins that belong to the swi2/snf2 helicase superfamily of ATPases and can use energy to slide nucleosomes along the DNA. They can be subdivided into the SWI/SNF, ISWI, CHD and INO80 families depending on the sequence homology of their ATPase domains (Table 1). While all remodelers have an affinity for nucleosomes, a variable domain and cofactor architecture directs their activity to specific loci and distinct processes by interaction with modified histone tails and other chromatin related proteins or transcription factors. The following paragraphs give a short overview of the main remodeler families and present some examples of their functions.

The SWI/SNF family of remodelers is an 8-12 subunit complex that slides and ejects nucleosomes in many diverse processes besides chromatin assembly. The catalytic ATPase contains a SANT domain followed by a bromodomain capable of binding to acetylated histone tails, and indeed histone acetylation increases the remodeling efficiency of SWI/SNF complexes and the enhanced RNA PolII elongation mediated by yeast RSC [38, 39]. SWI/SNF family members interact with multiple transcriptional activators (reviewed in [37]), and there is strong evidence for the SWI/SNF family to be involved in gene activation. dBrahma for instance is required for Pol II mediated expression of most genes on *Drosophila* polytene chromosomes [40], and yeast RSC lowers the nucleosome density at promoters, thereby promoting transcription [41].

ISWI family remodelers usually contain one or two catalytic subunits and a varying set of a few specialized attendant proteins. The ATPase contains a SANT followed by a SLIDE domain at its C-terminus, which together bind to an unmodified histone tail and DNA [42].

Family and composition		Organisms									
		Yeast			Fly			Human			
SWI/ SNF	Complex	SWI/SNF		RSC	BAP		PBAP		BAF		PBAF
	ATPase	Swi2/Snf2		Sth1	BRM/Brahma				hBRM or BRG1		BRG1
	Nucleocatalytic homologous subunits	Swi1/Adr6			OSA/eyelid				BAF250/hOSA1		
							Polybromo BAP170				BAF180 BAF200
		Swi3		Rsc8/Swh3	MOR/BAP155				BAF155, BAF170		
		Swp73		Rsc6	BAP60				BAF60a or b or c		
		Snf5		Sfh1	SNR1/BAP45				hSNF5/BAF47/INI1		
					BAP111/dalao				BAF57		
	Arp7, Arp9			BAP55 or BAP47				BAF53a or b			
			Actin				β-actin				
Unique	a		b								
ISWI	Complex	ISW1a	ISW1b	ISW2	NURF	CHRAC	ACF	NURF	CHRAC	ACF	
	ATPase	Isw1		Isw2	ISWI			SNF2L	SNF2H <sup>e</sup>		
	Nucleocatalytic homologous subunits			Ite1	NURF301	ACF1		BPTF	hACF1/WCRF180		
				NURF55/p55	CHRAC14		RbAp46 or 48	hCHRAC17			
					CHRAC16			hCHRAC15			
	Unique	Ioc3	Ioc2, Ioc4		NURF38						
CHD	Complex	CHD1			CHD1	Mi-2/NuRD		CHD1	NuRD		
	ATPase	Chd1			dCHD1	dMi-2		CHD1	Mi-2α/CHD3, Mi-2β/CHD4		
	Nucleocatalytic homologous subunits					dMBD2/3			MBD3		
						dMTA			MTA1,2,3		
						dRPD3			HDAC1,2		
						p55			RbAp46 or 48		
						p66/68			p66αβ		
Unique								DOC-1?			
INO80	Complex	INO80		SWR1	Pho-dINO80	Tip60		INO80	SRCAP	TRRAP/Tip60	
	ATPase	Ino80		Swr1	dIno80	Domino		hIno80	SRCAP	p400	
	Nucleocatalytic homologous subunits	Rvb1,2			Reptin, Pontin			RUVBL1,2/Tip49a,b			
		Arp5,8		Arp6	dArp5,8	BAP55		BAF53a			
		Arp4, Actin1			dActin1	Actin87E		Arp5,8	Arp6	Actin	
		Taf14		Yaf9		dGAS41		GAS41			
		Ies2,6		hIes2,6							
			Swc4/Eaf2	dDMAP1			DMAP1				
			Swc2/Vps72	dYL-1			YL-1				
			Bdf1	dBrd8							
			H2AZ,H2B	H2Av,H2B			H2AZ,H2B				
			Swc6/Vps71				ZnF-HIT1				
				dTra1			TRRAP				
				dTip60			Tip60				
				dMRG15			MRG15				
		dEaf6		MRGX							
		dMRGBP		FLJ11730							
		E(Pc)		MRGBP							
		dING3		EPC1, EPC-like							
		Unique	Ies1,Ies3-5,Nhp10			Swc3,5,7	Pho			d	

<sup>a</sup>Swp82, Taf14, Snf6, Snf11.

<sup>b</sup>Rsc1 or Rsc2, Rsc3-5, 7, 9, 10, 30, Htl1, Ldb7, Rtt102.

<sup>c</sup>In addition, SNF2H associates respectively with Tip5, RSF1, and WSTF to form **NoRC**, **RSF**, and **WICH** remodelers.

<sup>d</sup>Amida, NFRKB, MCRS1, UCH37, FLJ90652, FLJ20309.

**Table 1. Remodeler composition and orthologous subunits.** Taken from [37].



The ISWI family members ACF and CHRAC mostly optimize nucleosome spacing during chromatin assembly and transcriptional repression [42]. Yeast ISW2 for instance is recruited to many genes by the yUme6 repressor, and helps to generally repress cryptic transcription initiation [43, 44]. In contrast, the NURF complex can randomize spacing in support of transcription, emphasizing the functional diversity mediated by changes in subunit composition [45]. At the same time, the activity of NURF is required to maintain the chromatin structure of the male X chromosome in *Drosophila*, which undergoes selective, massive unfolding upon NURF inactivation [45, 46].

The CHD or Mi2 complex is monomeric in lower eukaryotes but can include up to 10 subunits in vertebrates. Characteristic are two tandemly arranged chromodomains at the N-terminus of the catalytic subunit, which in the case of human CHD can bind to the H3K4me2 or me3 marks correlated with active chromatin [47, 48]. In *S. pombe* CHD has been shown to eject nucleosomes at promoters to stimulate transcription activation [49]. All four CHD members colocalize with PolII on polytene chromosomes in *Drosophila* [50, 51], and yeast Chd1 interacts with multiple elongation factors suggesting a general role of CHD remodelers in Pol II elongation [51]. In contrast, the vertebrate Mi2/NuRD complex contains histone deacetylase activities and has repressive functions (reviewed in [52]).

Complexes of the INO80 family contain more than 10 subunits and are characterized by an insertion in the middle of the ATPase domain to which regulatory factors such as actin related proteins (ARP) can bind. INO80 can function in gene activation [53], and the related yeast Swr1 complex deposits the variant histone H2A.Z at gene promoters [54]. Furthermore, it has been shown in yeast and human that INO80 is required for the recovery of stalled replication forks during replicative stress in S-phase [55, 56]. In addition, INO80 has received a lot of attention for its role in DNA repair and recombination. INO80 and the complexes of the SWR1 subfamily are quickly recruited to the site of DNA double strand breaks where they help to create the 3' single-stranded DNA overhang critical for the repair mechanism [57].

The above examples for remodeler functions are far from being complete. The central role that energy dependent mobilization of nucleosomes plays for virtually all chromatin related processes has led to a huge diversification of this group of enzymatic complexes, and the limited space of this introduction cannot do full justice to the topic. The reader may therefore refer to some recent reviews for a more comprehensive introduction into the biology of chromatin remodeling complexes [37, 58].

We have seen so far how histones and DNA form chromatin in order to control accessibility of the DNA template, and how histone chaperones and chromatin remodeling

complexes together regulate the assembly of this structure. However, the resulting chromatin is far from being uniform. Instead, chromatin structure is further differentiated through posttranslational modifications of histones, and by the incorporation of alternative histone proteins, the so called histone variants.

### 1.1.6 Histone variants

Maybe the most direct way to modulate chromatin structure is the incorporation of non-canonical histone variants into the core nucleosome. Variant forms have been identified for all histones, and their origins date back to the earliest diversifications of eukaryotic lineages. Other than canonical histones, histone variants can be expressed in a replication dependent or independent manner and are usually expressed from one gene. The incorporation of variant histones can specifically alter the fundamental structure and stability of the core nucleosome itself, as well as the recruitment of chromatin modifying enzymes and transcription factors. Therefore, individual histone variants have evolved to serve specific processes such as transcriptional activation, repression or memory, DNA damage and centromere formation [3]. This chapter gives a short overview over the most common eukaryotic histone variants and points out some examples for their functions.

**CenH3**, also known as CENP-A according to the first identified mammalian example, functionally refers to the quickly evolving and relatively lowly conserved group of centromere specific H3. The histone fold domains of CenH3s show ~50-60% identity with canonical H3 and no conservation in the N-terminal tails. The mechanisms that maintain CenH3 at centromeres are not well understood, although CenH3 shows some preference for AT-rich DNA, which is usually enriched at centromeres [59]. The structure of CenH3 nucleosomes is unusual and controversial. In yeast and *Drosophila* CenH3 nucleosomes induce positive supercoils into the DNA, suggesting a right-handed helix [60]. Atomic force measurements show that *Drosophila* CenH3 (CID) nucleosomes are of half the height as canonical nucleosomes, and they protect less than the usual ~150bp of DNA against nucleases [61]. Accordingly, *Drosophila* CID nucleosomes contain one copy each of CID, H4, H2A and H2B and most likely resemble half of an octameric nucleosome, or a hemisome. The situation is not as clear in yeast, where either a hemisome or a (CenH3–H4)<sub>2</sub> tetrameric structure has been proposed for CenH3 containing nucleosomes.

**H3.3** in most animals only differs at 4 residues from canonical H3, including a change from Ala to Ser at position 31 in the N-terminal tail and substitutions at residues 87, 89, and

90 near the beginning of  $\alpha 2$ . H3.3 is incorporated and continuously exchanged in the body of transcribed genes during transcription elongation [62, 63], while the three substitutions in  $\alpha 2$  prevent replication independent incorporation of canonical H3 [64]. The turnover of H3.3 is generally higher as compared to H3, and this might contribute to the maintenance of active chromatin. Interestingly H3.3 is enriched on the male X chromosome in *Drosophila*, most likely reflecting the enhanced transcriptional activity resulting from dosage compensation [65]. An intriguing study using a *Xenopus laevis* nuclear transplantation assay furthermore showed that H3.3 can mediate epigenetic memory of the transcriptional state of a gene over 12 consecutive cell divisions, depending on the presence of Lys4 of H3.3 [66].

**H2A.Z** differs from H2A around the L1– $\alpha 2$  and  $\alpha 2$ –L2 junctions and in the C-terminal contact region to H3, rendering *in vitro* reconstituted nucleosomes more stable [67]. Of particular interest with respect to the formation of higher order chromatin structures is the more pronounced acidic character of the negatively charged “acidic patch” in H2A.Z, which lies at the flat surface of the H2A/H2B dimer [10, 68]. This region of H2A/H2B has been shown to bind to positively charged lysine residues in the H4 N-terminal tail of neighboring nucleosomes [69], and incorporation of H2A.Z might therefore favor inter-nucleosomal interactions. Nucleosomes containing H2A.Z are found on either side of the nucleosome depleted regions at gene promoters, where they are highly positioned relative to the TSS. Although H2A.Z promotes efficient recruitment of RNA PolII in yeast and human cells [70, 71], the roles of H2A.Z in gene regulation as well as other processes such as heterochromatin formation or DNA repair are contradictory, most likely due to variable posttranslational modifications and the effects of nucleosome remodeling complexes [72, 73]. One intriguing recent discovery was that H2A.Z is required for tethering of the INO1 gene to the nuclear membrane in yeast during the establishment of transcriptional memory [74].

**H2A.X** is characterized by a short C-terminal motif containing Ser139, which gets rapidly phosphorylated by ATM, ATR and DNA-PK kinases at nucleosomes surrounding DNA double strand breaks (DSB). The resulting phosphorylated form termed  $\gamma$ H2A.X helps to recruit DNA repair, chromatin remodeling and histone modifying enzymes to the site of DNA damage [73, 75]. However, the exact details of how H2A.X functions in DNA repair are still under active investigation. Beyond DNA repair, H2A.X is also found in numerous foci that are independent of DSBs and whose significance is still unclear. Furthermore,  $\gamma$ H2A.X is also required for sex chromosome inactivation in the male germline and for the silencing of unpaired meiotic chromatin [76, 77].

**MacroH2A** is an H2A variant specific to animals. It is characterized by an additional “macrodomain” at its C-terminus, which is able to recognize ADP-ribose metabolites. On reconstituted chromatin mH2A restricts transcription factor access, probably due to sterical hindrance by the relatively large macrodomain, and prevents gene activation by p300 [78]. In a human pluripotent cell line, mH2A correlates with genes that are repressed by the polycomb system and carry H3 Lys27 trimethylation, consistent with a role of mH2A in conditional gene silencing [79], and a suppressive role of mH2A is further suggested by its enrichment on the human inactive female X chromosome [80]. In addition to its functions in gene repression mH2A has recently been shown to be recruited to activated poly-ADP-ribose polymerase at sites of DNA damage, where it helps to reorganize the chromatin surrounding the lesion [81].

Taken together, histone variants allow local reprogramming of the chromatin environment, in order to perform spatially or temporally restricted tasks. This concept is dramatically extended by the introduction of posttranslational modifications that further differentiate the properties of histone proteins.

### 1.1.7 Posttranslational modification of histones

It has been known for nearly 50 years that acetylation of histones is associated with active genes *in vivo* [82]; however, it remained unclear if acetylation was cause or merely consequence of transcription. With the identification of a histone acetyltransferase activity in a ciliate homologue of the known yeast coactivator Gcn5, a first functional link between gene activation and the covalent modification of histones became apparent [83]. Subsequently, other previously known coactivators like CBP/p300, SRC-1, and ACTR were shown to be associated with histone acetyltransferase activities [84-87], and simultaneously the known transcriptional repressor Rpd3 was identified as a histone deacetylase [88]. Since then, more than 60 different often reversible posttranslational modifications of histones, including acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, ADP-ribosylation, deimination and proline-isomerization, have been identified, many of which can physically alter the chromatin structure or impact on the recruitment of chromatin modifying factors, thus creating a combinatorial potential of astonishing complexity (Figure 2, reviewed in [89, 90]). As a consequence, the old inherently static model of gene activation, in which sequence specific transcription factors recruit the transcriptional machinery to their target genes had to be replaced by a much more dynamic view, where spatial and temporal combinations of



acetylation of Lys16 prevents the formation of a 30nm fiber *in vitro* [98]. In male flies, the majority of X chromosomal chromatin is hyper-acetylated at H4K16, leading to chromatin decompaction and higher accessibility for the transcription machinery. *Drosophila* dosage compensation therefore offers an instructive model system for the establishment of a global chromatin environment by histone modifications.

The second and most characterized mode of operation for histone modifications is the creation or occlusion of docking sites for chromatin binding proteins. A variety of adaptor domains has evolved to specifically recognize certain histone modifications and recruit chromatin modifying factors. For example the H3K9me3 mark enriched at constitutive telomeric and pericentric heterochromatin is recognized by a chromodomain in the HP1 protein [99, 100], which in turn recruits further histone deacetylase and methylase activities that lead to chromatin compaction [101-103]. As one of the first described examples of crosstalk between different histone modifications, the binding of HP1 to H3K9me3 can be disrupted by an adjacent phosphorylation of histone H3 at S10 [104]. Importantly, HP1 recruits Su(var)3-9, the enzyme responsible for the H3K9me3 mark itself, and it has been proposed that this allows the transmission of the mark onto the newly synthesized daughter strand during replication to implement epigenetic memory.

Another modification associated with the repressed state of genes is H3K27me3, mediated by the polycomb repressive complex 2 (PRC2) [105]. Initially described as a suppressor of Hox gene expression in *Drosophila* [106], hundreds of developmentally regulated genes have been identified as polycomb targets in the recent years [107]. In the fly, the initial recruitment of PRC2 to its target genes involves the recognition of so called polycomb response elements (PREs) by DNA binding factors (reviewed in [108]). Subsequent methylation of H3K27 then contributes to the recruitment of PRC1, which eventually leads to gene repression by an as yet unknown mechanism that involves monoubiquitination of H2A at K119.

The coordinated recruitment of a series of chromatin modifying proteins by consecutively established histone marks is a common theme that allows the temporal coordination of multistep processes. Another prominent example for the complex interplay of histone marks and the chromatin modifying machinery is the regulation by Paf1 and the COMPASS/TRX/MLL complexes in yeast, *Drosophila* and mammals, respectively. Upon initial recruitment to the promoter together with RNA PolII, Paf1 serves as a recruitment platform for COMPASS, which contains Set1, the only H3K4 methyltransferase in yeast [109]. At the same time, Paf1 stimulates the independently recruited rad6-Bre1 E1/E3 ligase

complex to monoubiquitinate H2B at K120, and only the presence of this histone mark triggers the full potential of Set1 to establish H3K4me3 [109, 110]. Finally, depending on the context, H3K4me3 can be read by PHD fingers, chromodomains, TUDOR and MBT domains of a variety of factors including chromatin remodelers, histone acetyltransferase complexes and TFIID, mostly in support of transcription [89, 111, 112].

Following the original observation that certain histone modifications like acetylation of H3 and H4 or methylation of H3K4 correlate with actively transcribed regions, while others such as methylated H3K9 and H3K27 are enriched in repressed chromatin, the “histone code hypothesis” was proposed, stating that certain combinations of histone marks will allow the prediction of distinct transcriptional outputs [113]. However, this simplistic view has been increasingly challenged in recent years, when more and more examples for a context specific, often antipodal read-out of histone marks were discovered. H3K9me3 for instance was found in the coding region of transcribed genes, where it closely follows the pattern of H3K36me3 [114]. Similarly, certain adaptors that recognize specific histone modifications have been found in activating as well as repressive complexes. As an example, yeast eaf3, which binds to H3K36me3 and directs the deacetylase activity of the Rpd3S complex to the body of transcribed genes in order to prevent cryptic initiation, is at the same time found in the NuA4 acetyltransferase complex [115-118]. Likewise, many histone marks can recruit opposing activities. In addition to the above mentioned positively acting factors, H3K4me3 is also recognized by demethylase activities or by the PHD finger containing ING2 protein in the mSin3a-HDAC1 histone deacetylase complex, which shuts off transcription of proliferation specific genes upon DNA damage [119, 120].

A very instructive example for the spatial and temporal orchestration of histone marks is a recent study of Zippo et al. that demonstrates how a cascade of histone modifications can lead to a particular transcriptional outcome in a context dependent manner [121]. The authors found previously that serum activation of the FOSL1 gene requires the PIM1 kinase, which is able to phosphorylate H3S10 [122]. MSK1 and MSK2 also mediate H3S10 phosphorylation at the FOSL1 promoter at early time points of expression [123, 124]. In their new study Zippo and colleagues show, that phosphorylation of H3S10 by PIM1 at the FOSL1 enhancer is recognized by the 14-3-3 protein, which in turn recruits the MOF histone acetyltransferase to the enhancer. The resulting acetylation of H4K16 at the FOSL1 promoter forms a binding site for Brd4, a component of the P-TEFb kinase, which phosphorylates the C-terminal domain of PolII to activate transcription. In contrast, earlier recruitment of 14-3-3 by H3S10p at the FOSL1 promoter does not lead to MOF recruitment, thereby demonstrating the importance of

the spatiotemporal context for the determination of downstream events. As mentioned above, H4K16ac is also the hallmark of the hyperactive male X chromosome in *Drosophila*. Interestingly, the H3S10 specific Jil1 kinase is colocalizing with MOF on the male X. However, the recruitment of Jil1 occurs later than hyper-acetylation of H4K16 [125], suggesting a biological function that is unrelated to the one observed at the FOSL1 gene.

All different means of chromatin regulation introduced so far are, alone or in combination, effective tools to modulate local chromatin structure. It is important to recognize however, that cells represent sophisticated global networks of gene expression, where the activity of a gene often feeds back into the regulation of other genes [126]. Therefore, to maintain homeostasis of the transcriptome as a whole, the regulation of transcription and chromatin structure has to be an integrated process that allows the concerted regulation of spatially distant regions of the genome. It has become increasingly clear in recent years that non coding RNA transcripts play a major role in orchestrating this global chromatin regulation, as I will explain in more detail in the next section.

### **1.1.8 Noncoding RNAs and chromatin**

**Small RNAs**      Already with the discovery of the RNAi pathway in the 1990s, the role of RNA as a mere transmitter of information along the central dogma 'DNA makes RNA makes protein' was put into question [127]. Many short regulatory RNAs were subsequently discovered that have roles in transcriptional control and chromatin organization (reviewed in [128]). As a common theme for regulation by small RNAs, a double stranded RNA sequence is processed by the endonuclease Dicer into a ~21nt siRNA, which is then bound by a protein of the Argonaute family, thereby forming the RNA induced transcriptional silencing complex (RITS). Argonaute bound siRNAs are then often amplified by an RNA dependent RNA polymerase (RdRP). Together with Argonaute, the siRNAs subsequently find and inactivate target mRNAs by perfect base pairing, followed by degradation of the target.

RITS complexes can also bind to nascent transcripts and recruit chromatin modifying enzymes. This mechanism is exemplified in *S. pombe*, where low levels of transcribed RNA from centromeric repeats are processed by Dicer, and Argonaute directed recruitment of the H3K9 methyltransferase Clr4 to the site of transcription then leads to heterochromatin formation [128]. Interestingly, RNA also plays a role in kinetochore integrity during mitosis in human cells [129, 130], and the formation of human neocentromeres depends on Dicer [131], although the molecular steps involved have not been identified. In addition to silencing



of pericentric heterochromatin, siRNAs mediate the silencing of transposable DNA elements in plants by similar mechanisms [128], and have been shown to play a role for transposon silencing in somatic tissues of *Drosophila* [132-134]. In the germ line of *Drosophila* and vertebrates, transposon silencing is mediated by piRNAs that are not dependent on Dicer for their biogenesis and interact with the Piwi clade of Argonaute proteins (reviewed in [135]).

In addition, a new class of small ~50 - 200bp RNAs that function independently from the siRNA pathway has recently been identified as part of the polycomb repression system. Short RNAs transcribed from the 5' end of repressed genes in primary human CD4+ T cells and murine ES cells form hairpin structures that are recognized by Suz12 in the PRC2 complex, and 5' fusion of such a hairpin sequence repressed transcription of a reporter gene [136].

### **lincRNAs**

In the recent years, the development of whole-genome tiling-arrays and high-throughput sequencing technologies has led to the discovery that the majority of the genome in higher organisms is transcribed into RNA, including a new class of long intergenic noncoding RNAs (lincRNA) larger than 200bp (reviewed in [137]). Initially regarded as transcriptional noise [138], it rapidly became clear that lincRNAs are produced in a developmentally regulated fashion and under positive evolutionary selection [139-142]. In the meanwhile, mounting evidence suggests that this vast hidden layer of lincRNAs plays a major role in the epigenetic regulation of multicellular organisms.

lincRNAs can interact with chromatin and the associated factors in multiple ways to exert their functions. One of the first described examples was the X chromosome linked Xist RNA, whose expression is required for somatic silencing of one female X chromosome in mammals during dosage compensation, a process known as X inactivation [143]. By now, a complex interplay of partially overlapping and antisense RNAs has been identified that also includes RepA and Tsix. The Xist antagonist Tsix and RepA, which also contains sequences from the 5' end of Xist, compete for binding to PRC2. Upon downregulation of Tsix on the future inactive X chromosome (Xi), RepA can efficiently recruit PRC2 to the Xi and activate expression of Xist. Xist then distributes PRC2 along the Xi, leading to H3K27me3, H2AK119Ub and ultimately DNA methylation and compaction of the Xi into the barr body ([144] and reviewed in [145]). Analogous to the mammalian situation, yet with an inverse outcome, roX (RNA on the X) RNAs are required for the transcriptional activation of the single male X chromosome during *Drosophila* dosage compensation, which will be discussed in more detail below.

The recruitment of chromatin remodeling activities by lincRNAs turned out to be a widespread and general phenomenon. Another prominent example is the process of genomic imprinting in marsupials and eutherian species, during which genes get permanently repressed in the germ line of either the male or female sex. Here it has been shown that the 108 kb Air RNA tethers the H3K9 specific methyltransferase G9a to the Slc22a3 promoter during male specific silencing of the cis-linked Slc22a3, Slc22a2, and Igf2r genes in the placenta [146]. Similarly, the 90.5 kb long Kcnq1ot1 RNA mediates silencing of 10 paternally imprinted genes in the Kcnq1 domain by recruiting G9a and PRC2 [147]. Regulation by lincRNAs can also function in *trans*, as exemplified by the 2.2 kb HOTAIR, which is transcribed from the HOXC locus and mediates silencing across 40kb of the HOXD locus on another chromosome. HOTAIR is not only bound by Suz12 and Ezh2 in the PRC2 complex, but also recruits the LSD1 lysine demethylase, leading to coupled histone H3 lysine 27 methylation and lysine 4 demethylation at the HOXD locus [148]. Another example involving recruitment of PRC2 is the repression of p53 target genes by the lincRNA-p21 upon activation of the p53 pathway [149]. In the meantime, thousands of PRC2 interacting RNAs have been identified, which are likely to orchestrate PRC2 function across the mammalian genome [150].

To further increase the complexity of this new layer of genomic regulation, lincRNAs can exert their biological functions by a variety of additional mechanisms. They can recruit RNA binding proteins to modulate promoter function, like in the case of the cyclin D1 promoter, where upon DNA damage lincRNAs recruit the TLS protein to inhibit the acetyltransferase activity of CREB and p300 and silence cyclin D1 expression [151]. They can also act as cofactors to promote transcription, as in the case of the Dlx6 gene in mice, where the lincRNA Evf2 recruits and activates the transcription factor DLX2 at an ultraconserved enhancer [152]. lincRNAs can directly control promoter choice by RNA PolII, for instance at the dihydrofolate reductase (DHFR) locus, where a lincRNA transcribed from an upstream locus can form a RNA/DNA triplex in the major promoter of DHFR that prevents recognition by TFIID [153]. Finally, lincRNAs can directly interact with basal components of the transcriptional machinery to modulate PolII function. For example, Alu elements that are transcribed by PolIII in humans upon heat shock can bind tightly to PolII to prevent the formation of active preinitiation complexes [154].

A variety of other functions has been described for lincRNAs, involving diverse biological processes like mRNA splicing or the organization of nuclear speckles ([155], and reviewed in [156]), which are outside the scope of this introduction to chromatin, but again emphasize the diversity and regulatory potential of this relatively new class of molecules.

### 1.1.9 Functional states of chromatin

The combined action of all the above mentioned mechanisms not only allows for the regulation of individual genes in response to environmental cues and developmental programs. They also enable the cell to establish functional chromatin states across extended genomic regions or even whole chromosomal territories. These functional states can vary in extent and molecular composition, and the consequences for transcription can range from hyperactivation, like on the male X chromosome in *Drosophila* dosage compensation, to gene repression as in polycomb mediated repression of HOX gene clusters during cell differentiation.

In the oldest classification of distinct chromatin domains, well before the discovery of DNA as the carrier of the genetic information, heterochromatin and euchromatin had been defined as regions inside the nucleus, according to their susceptibility to staining with basic dyes [157]. Remarkably, until today, this very early classification has remained a useful criterion for the functional distinction of chromatin states. In general, the more condensed and closed architecture of heterochromatin renders it relatively inaccessible for transcription factors, resulting in low transcriptional activity. In contrast, euchromatin correlates with an open chromatin architecture; and high density gene clusters are found preferentially in open, euchromatic fibers [158]. However, there is no strict correlation with the gene expression status, since inactive genes can be found in euchromatic regions and some heterochromatic genes show active transcription [159]. An important distinction is made between constitutive and facultative heterochromatin. Constitutive heterochromatin mainly contains repetitive, gene-poor, and late replicating DNA sequences, such as transposable elements, telomeres and centromeres, is enriched for H3K9me2/3 and HP1 and is always compacted. In contrast, facultative heterochromatin can be actively condensed or decondensed, in order to shut off or activate embedded genes as part of developmental programs. The establishment of facultative heterochromatin is regulated by the polycomb/trithorax system and correlates with H3K27me3 in the repressed state. One interesting feature of heterochromatin is its ability to spread into neighboring chromatin regions. If a reporter gene is inserted into, or in close proximity to heterochromatin, its transcription can frequently become shut off, a well studied phenomenon known as position effect variegation (PEV). PEV has proven an invaluable tool for the identification of Su(var) and E(var) genes that suppress or enhance the formation and spreading of heterochromatin, respectively [160].

Very recently, several studies have suggested finer classifications of chromatin states, based on the comprehensive genome wide analysis of transcripts, histone modifications, chromosomal proteins, transcription factors, and nucleosome properties across developmental time courses and in multiple cell lines [141, 142, 161, 162]. It is now clear that functional states of chromatin cannot be predicted by the simple presence of a certain histone modification or non-histone protein. Instead, they are the integrated result of a vast number of protein factors, different RNA species and histone marks, which together orchestrate the formation of various chromatin environments with distinct properties. This concept becomes especially apparent during the process of X chromosome compensation in *Drosophila*, as I will exemplify in the final part of this introduction.

## **1.2 Dosage compensation in *Drosophila***

### **1.2.1 Sex chromosomes and the need for dosage compensation**

Complex genomes represent sophisticated networks, in which the expression of a gene is the result of the activity of many other genes [126]. Loss or gain of gene copy number has therefore the potential to disrupt the homeostasis of this intricate system, and genomic deletions are generally lethal in *Drosophila* when they include more than ~3% of the genome [163]. However, reoccurring copy number variation is not always the consequence of deleterious processes. In many organisms, sex determination involves an unequal distribution of sex chromosomes, leading to chromosomal aneuploidy in the heterogametic sex. The most common forms are the XX/XY system and the ZZ/ZW system, with males or females being the heterogametic sex, respectively (reviewed in [164, 165]). In a generally accepted model, the initial trigger for the evolution of sex chromosomes from a previously identical set of autosomes is the occurrence of a male-determining gene, which is only propagated in males. To be effective, recombination of this gene needs to be suppressed, creating a neo-Y chromosome that is unable to recombine with the corresponding sister chromosome, the future neo-X. Lack of recombination leads to progressive degeneration of the Y chromosome, rendering the male sex hemizygous for the genes encoded on the remaining X chromosome. To prevent the deleterious effect of this loss of gene dose, flies have evolved a dosage compensation system that upregulates transcription from the single male X chromosome by twofold (Figure 3, reviewed in [166]). Interestingly, all potential solutions to this problem have been realized throughout evolution. In humans, one female X chromosome is completely

shut off to equilibrate expression between the sexes, while in *C. elegans* (where the male sex has an XO genotype) expression from both X chromosomes in hermaphrodites is repressed by half [167, 168]. In mammals and *C. elegans*, X chromosome repression in the homogametic sex would lead to an imbalance between X-linked and autosomal gene expression. Therefore, a second mechanism is thought to be active in these organisms that upregulates X-linked transcription in both sexes by an as yet unknown mechanism [169, 170]. Curiously, recent work in several bird and lepidopteran species of the female heterogametic ZZ/ZW type suggested that not all organisms have the same requirement to fully compensate transcription from unequally distributed sex chromosomes [171].

	♀	♂
<i>Drosophila</i>	XX AA	X <sup>2</sup> Y AA
Mammals	X <sub>x</sub> AA	X <sup>2</sup> Y AA
<i>C. elegans</i>	XX AA	X AA

**Figure 3: Dosage compensation strategies. From [172].**

In *Drosophila*, transcription from the single male X chromosome is upregulated by twofold. In mammals, transcription from one female X chromosome is silenced during X inactivation. In *C. elegans*, transcription from both female X chromosomes is repressed by half. It is believed that in mammals and *C. elegans*, X-linked transcription is generally activated by an independent mechanism to match autosomal expression levels. Upregulated chromosomes are in green; chromosomes subject to transcriptional enhancement and repression are in yellow; the letter ‘A’ denotes autosomes.

### 1.2.2 The MSL complex

*Drosophila* dosage compensation is one of the oldest model systems for the study of chromatin and transcription regulation. Already in 1922 it was observed that hemizygous males have the same levels of eye pigmentation as heterozygous females [173]; and the term “dosage compensation” was later introduced based on the same phenomenon [174]. These observations were subsequently confirmed biochemically for enzymes encoded on the X chromosome [175]; and enhanced incorporation of H3 uridine into RNA produced along the male X chromosome finally suggested that dosage compensation is achieved at the

transcriptional level [176]. The development of increasingly sophisticated genetic tools eventually led to the identification of the factors responsible for this transcriptional activation. After the first mutant allele of the male specific lethal gene MLE had been extracted from a natural population of *D. melanogaster* in Kofu, Japan [177], several genetic screens aimed at the identification of more genes with male specific phenotypes. These efforts soon lead to the identification of the male specific lethal 1 (MSL1), MSL2 and MSL3 proteins [178]. All MSLs were subsequently found to colocalize at hundreds of sites along the male X chromosome, suggesting a direct involvement in the transcriptional activation of X chromosomal genes [179-181]. MSL binding to X-linked chromatin was shown to be dependent on the MSL2 protein, which is not present in females due to translational repression by the master sex regulator sex-lethal [181-183], thereby ensuring male specificity of the dosage compensation process. Consistent with the idea that MSLs can directly modulate chromatin structure, it was also discovered that the male X chromosome in *Drosophila* is marked by high levels of H4K16ac, and that this hyper-acetylation is dependent on MSL function [179, 184]. Accordingly, a fifth MSL protein, males absent on the first (MOF), was identified soon after and shown to be a H4K16 specific histone acetyltransferase [185-187]. A model had emerged in which MSLs together bind to the male X chromosome as part of the dosage compensation complex (DCC), leading to hyper-acetylation and thus transcriptional activation of X-linked genes [188]. The DCC eventually turned out to be a ribonucleoprotein complex, when two additional functionally redundant non-coding RNAs (RNA on the X, roX1 and roX2) were identified as further essential components [189-192]. Interestingly, MLE was revealed in the meanwhile as the homologue of human RNA helicase A, and its ATPase activity is required for X chromosome targeting [193, 194]. Upon treatment with RNase A, MLE is specifically removed from the male X chromosome [195]; and the delocalization of other MSLs observed in MLE mutants is very similar to the phenotypes resulting from the absence of both roX RNAs [191, 194].

### 1.2.3 Programming of the male X chromosome

Both roX RNA genes share another very interesting feature. Not only are their gene products components of the DCC, both genes are also located on the X chromosome and are themselves high affinity targets for DCC binding [196]. One early observation during the study of MSLs was the fact that, while MSL1 and MSL2 are absolutely essential for X chromosome targeting, core MSL complexes can bind to a subset of target loci throughout the

X chromosome even in the absence of MLE, MSL3 or MOF [180, 185, 197]. It was then shown that the roX genes retain their ability to nucleate MSL binding when moved to an autosomal location. Based on these observations a model was proposed in which the DCC would assemble with roX RNAs directly at their sites of transcription and then “spread” to other high affinity sites (HAS) on the X chromosome, and eventually to lower affinity binding sites [196]. Early attempts to define the nature of HAS in more detail, or to define consensus DNA sequences for MSL binding, were limited by the resolution of immunostainings and by the lack of more appropriate genome wide methods [198]. These limitations have been overcome in recent years by the development of whole genome microarrays and deep sequencing technologies. Coupled to chromatin immunoprecipitation (ChIP), ChIP-chip and ChIP-Sequencing methods have allowed to determine the HAS in high resolution, and to define corresponding DNA sequence motifs [199, 200]. At the same time, the precise MSL binding patterns on dosage compensated genes could be obtained, showing that MSLs are preferentially bound to the transcribed regions and 3’ ends of their target genes [201-203]. In addition, a series of targeting cues for the recruitment of MSLs to their low affinity targets have been identified recently, and we now begin to understand the formation of the dosage compensated X chromosome as a sophisticated multistep process. Using immuno-FISH and confocal microscopy it was shown that HAS cluster together upon MSL binding, thereby forming a three dimensional X chromosomal territory [204]. roX RNAs are expressed from within this territory, and the immediate assembly with MSL1 and MSL2 alters the chromatin binding properties of these proteins [205]. In combination with the local clustering of high affinity MSL binding sites, this leads to the retention of MSLs, forming an environment of high local MSL concentration [206]. It has been known for some time that spreading of the DCC to target sites of lower affinity throughout the X chromosome is dependent on overall MSL protein levels [207]. Indeed, the high MSL concentration within the X chromosomal territory allows the subsequent recognition of a combination of targeting cues in the transcribed region of X-linked genes, most of which would not be targeted by the DCC in any other chromosomal context. It has been shown using X to autosome translocations that these signals include the transcription dependent recognition of degenerate DNA sequences in the 3’ end of dosage compensated genes [208]. Accordingly, the dependency of MSL spreading on the catalytic activity of MOF might be explained by the need to expose these signals through H4K16ac mediated relaxation of the chromatin structure [185]. However, DNA sequences are clearly not the only signal for MSL recruitment into the body of X linked genes, since autosomal genes can get recognized and compensated by the DCC when inserted

into the X chromosome, provided that they are actively transcribed [209]. Accordingly, transcription coupled histone modifications have been identified as additional signals for MSL recruitment. Soon after the finding that *eaf3*, the yeast homologue of MSL3, can bind to trimethylated H3K36 via its chromodomain [116, 117], it was shown that MSL binding highly correlates with the pattern of H3K36me3 on the male X chromosome in *Drosophila*, and that removal of the enzyme responsible for this histone mark leads to a loss of MSL spreading into X-linked genes [210, 211]. In line with these results, a similar phenotype was observed upon deletion of the MSL3 chromobarrel domain [212]. Interestingly however, no direct interaction could be detected between the MSL3 chromobarrel domain and an H3 peptide trimethylated at K36. Instead the MSL3 chromobarrel domain was found to crystallize together with the N-terminal tail of histone H4 monomethylated at K20, suggesting a potential role of this mark for MSL function [213, 214]. Clearly more work is required to reveal all targeting cues that help to specifically direct the MSL complex to dosage compensated genes. However, the recent advances in our understanding have unraveled the general principle that a combination of transcription coupled signals is specifically recognized when located inside the especially programmed environment of the male X chromosome.

#### **1.2.4 Transcriptional activation**

The final step in this chain of events is the eventual upregulation of X-linked transcription. As discussed before, one hallmark of the dosage compensated X chromosome is the MOF mediated hyper-acetylation of H4K16, the only histone mark that directly affects chromatin compaction through the disruption of inter-nucleosomal contacts [98, 215]. Furthermore, it has been shown that tethering of MOF leads to transcriptional activation of a reporter gene in yeast [187], and H4K16ac is correlated with high transcription in S2 and KC cell lines [216]. It is therefore believed that the permissive chromatin structure associated with H4K16ac leads to increased transcription factor accessibility, less energy consuming progression of PolIII through the chromatin template, and consequently transcriptional upregulation of the male X chromosome. However, final prove that H4K16ac is the direct cause of the transcriptional activation is still missing. In an alternative model, the mark would just serve to open up chromatin for the binding of other MSLs, which would then upregulate transcription using an independent mechanism. Also at which step in the transcription cycle dosage compensation is effective is still unknown. A recent study showed higher overall levels of PolIII on the X chromosome in S2 cells and attributed this to enhanced release into



the transcribed regions of X-linked genes as well as higher PolII processivity [217]. However, the authors did not compare this pattern of PolII to the female situation, leaving the possibility that differences in PolII levels at gene promoters might exist between sexes.

### 1.2.5 Other factors

Besides MSL proteins themselves, a number of other factors have been found associated with dosage compensation in *Drosophila*. The Jil-1 kinase phosphorylates serine 10 of histone H3 and is found enriched on the male X chromosome [218]. The H3S10ph mark is thought to stabilize active chromatin conformations, but depletion of Jil-1 has only moderate effects on X-linked transcription [219]. Furthermore, normal levels of the heterochromatin protein Su(var)3-7, a known interactor of HP1, are required to maintain X chromosome structure [220, 221]; and massive bloating of the male X chromosome has been observed when the nucleosome remodeling factor (NURF) is inactivated [45, 46]. Conversely, bloating of the male X chromosome has been observed upon overexpression of the DNA supercoiling factor (SCF), which is also required for the compensation of X-linked genes [222]. In addition, it has been shown that wt levels of Upstream of N-ras (UNR), a necessary cofactor for translational repression of MSL2 in females, are critical for X chromosome targeting of the DCC in males [223]. Unexpectedly, also the nuclear pore components Nup153 and Mtor have been found associated with the MSL complex [224]. Both factors are required for targeting of the MSL complex and bind to extended regions on the male X chromosome [216]. It has been hypothesized that tethering to the nuclear pore might facilitate efficient export of X-linked transcripts from the nucleus, and both nucleoporins appear generally associated with actively transcribed region across the *Drosophila* genome [216]. To make the picture even more complex, recent work has suggested that genomic deletions and duplications on all chromosomes are transcriptionally compensated in *Drosophila*, and that the twofold activation we observe on the male X chromosome is the combined result of global as well as X specific compensation mechanisms [225, 226]. Analogue to the use of multiple targeting signals for DCC recruitment, it is conceivable that the combination of multiple compensation systems will result in the most robust transcriptional output, while ensuring flexibility at the individual gene level.

### 1.2.6 Integration of MOF into the dosage compensation complex

Interestingly, all MSL proteins besides MSL2 are also present in females, and we are only beginning to understand their role in transcription regulation outside the context of the male X chromosome. In addition to its role in dosage compensation, MOF has recently been found at hundreds of active promoters across the whole genome in male and female cell lines, where it is bound as part of the newly identified Non-Specific Lethal (NSL) complex [215, 224, 227]. While several studies have found H4K16ac at the 5' end of genes in *Drosophila* [215, 228, 229], it has remained unclear if MOF is active as a HAT at these sites, or if it has indeed any chromatin related function outside the context of the male X chromosome [230]. Furthermore, it is not known how MOF targeting and activity are differentially regulated and distributed between the NSL complex and the DCC. MOF contains a zinc finger in its C-terminal HAT domain that is characteristic for the MYST family of acetyltransferases. In the case of *Drosophila*, this region has been shown to integrate MOF into the DCC by interaction with the PEHE domain of MSL1 [231]. Recently, a similar PEHE domain in NSL1, a member of the NSL complex, has been shown to interact at the same region [232]. Disruption of the MOF zinc finger furthermore prevents nucleosome binding of MOF *in vitro* [233]. Within the MSL complex there are two proteins that show similarities to chromo-like domains, MSL3 and MOF. As mentioned before, the chromobarrel domain of *Drosophila* MSL3 has been shown to interact with DNA and monomethylated H4K20 [213, 214]. Accordingly, mutations in the MSL3 chromobarrel domain lead to reduced male viability, most likely due to defects in dosage compensation [213, 214, 234]. MOF also contains a chromobarrel domain, which in this case lacks the aromatic cage necessary for methyl lysine binding [235]. Instead, the MOF chromobarrel domain is required for MOF binding to roX RNAs *in vitro* and *in vivo* [236]. Remarkably however, ten years after this initial observation, the function and biological significance of the MOF chromobarrel domain have remained elusive. In fact, no biological role has been described for chromobarrel domains of the MOF-type in general, which are conserved in histone acetyltransferases from yeast to human [237]. In the absence of both roX RNAs or the RNA helicase MLE, MSLs are delocalized from the X chromosome and ectopically targeted to autosomal sites, the chromocenter and the 4<sup>th</sup> chromosome [185, 191]. In line with these results, MOF is delocalized from the X chromosome upon treatment of permeabilized nuclei with RNase A [236]. Furthermore, deletion of a short "roX box" in roX2 leads to a specific reduction of H4K16ac on the male X chromosome, while MOF targeting itself remains unaffected [238, 239]. However, in marked contrast to the above

results, an earlier report suggested that the chromobarrel domain is dispensable for MOF function [231]. The remaining N-terminal half of the MOF protein appears inherently unstructured and has no functions assigned to it. Interestingly, although MOF has orthologs from yeast to mammals comprising a chromobarrel domain followed by a MYST-type HAT domain, the presence of a long N-terminal region is specific to *Drosophila* species. It seems therefore plausible that this domain is an evolutionarily late acquirement of *Drosophila* and involved in MOF functions that are specific to flies.



## **Aims and Objectives**

## 2. Aims and Objectives

In the study presented in chapters 3 and 4, I set out to investigate the full extent of MOF functions inside and outside the context of the male X chromosome. I was furthermore interested to recapitulate how various domains in the MOF protein differentially regulate these functions in order to utilize this conserved HAT for dosage compensation in *Drosophila* species. I've been able to show that MOF is responsible for genome wide H4K16ac in male and female flies. Using ChIP-Sequencing (ChIP-Seq) we find that the vast majority of all active genes, X-linked as well as autosomal, are MOF targets in both sexes. MOF function is essential in females, since loss of MOF leads to a severe reduction in female number and lifespan. I next show that the *Drosophila* specific N-terminus of MOF controls assembly of the DCC on the male X chromosome, and the HAT activity of MOF. Therefore, it seems plausible that the addition of this region to the MOF protein in *Drosophila*, resulting in a twofold increased protein size compared to mammalian MOF, has allowed to diversify MOF functions and to utilize it for dosage compensation. In addition, I've been able to demonstrate that the MOF chromobarrel domain serves to trigger H4K16ac after MOF binding to chromatin *in vivo* and *in vitro*. Disruption of the chromobarrel domain leads to a genomewide loss of H4K16ac and consequently compromised MSL spreading into X-linked genes. Multiple levels of control therefore reflect the enhanced complexity of MOF functions in flies and the resulting need for increased context dependent regulation of MOF activity. All bioinformatics analysis involved in this work was performed by Florence Cavalli at the European Bioinformatics Institute in Hinxton. The main findings presented in chapter 4 are currently under peer review for publication. The submitted manuscript, which was written by myself, can be found in Appendix C.

In chapter 5, I will describe ongoing work that aims at elucidating the mechanism of transcriptional upregulation of male X-linked genes during dosage compensation in *Drosophila*. To this end, I generated genomewide RNA PolIII ChIP-Seq profiles from 3<sup>rd</sup> instar larva salivary glands from male, female and male MSL2 RNAi flies. Analysis of this data suggests that dosage compensation operates at the level of transcription initiation. This result constitutes a major advance in the field of dosage compensation and is currently being verified by additional control experiments.

In chapter 6, I present my contributions to two major publications, in which I investigated the role of the H3K36 specific methyltransferase HypB/Set2 for dosage

compensation and the role of MOF for NSL function in *Drosophila*. Both original manuscripts can be found in Appendix A and B.

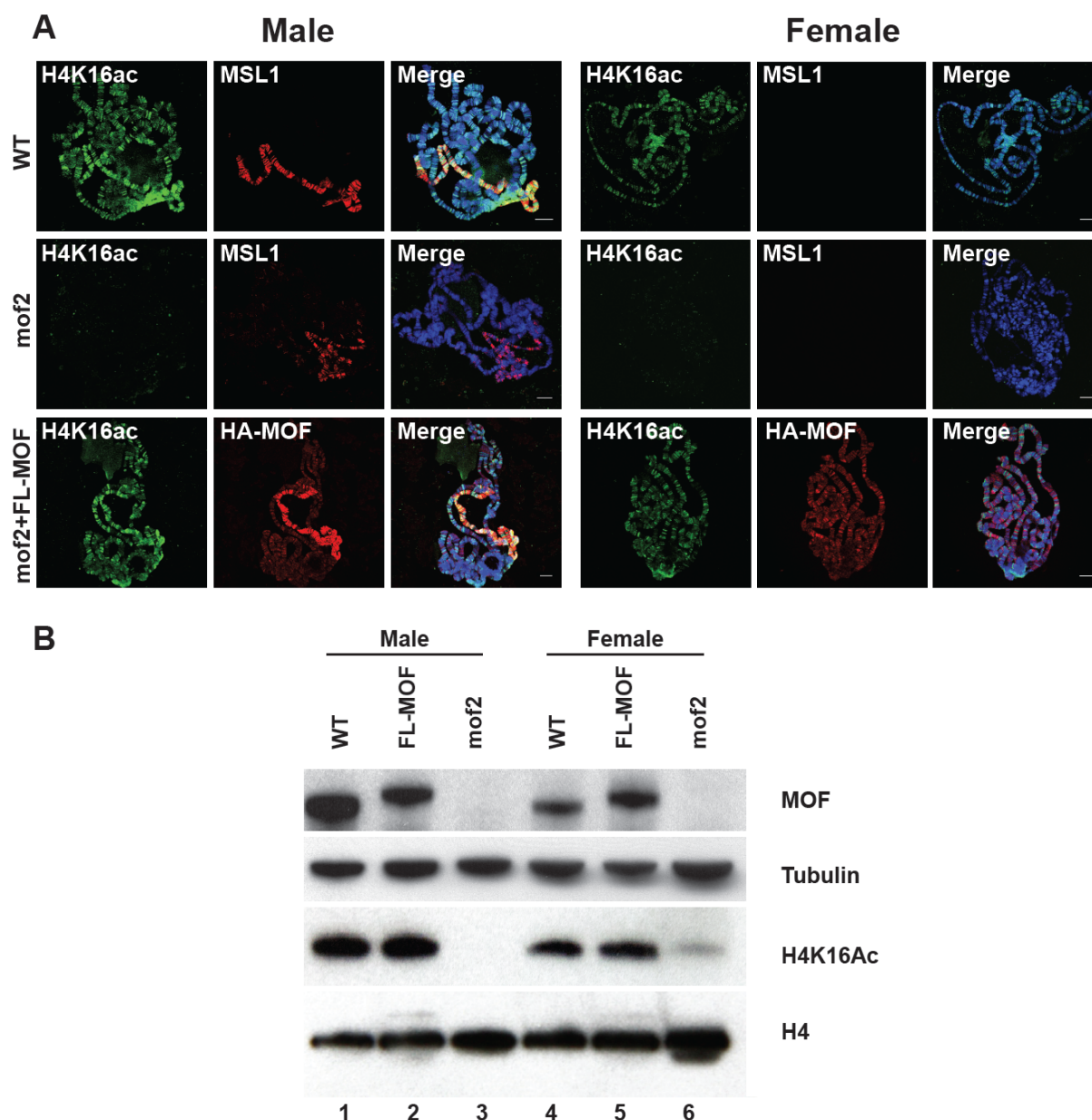




# Results

### 3.1 MOF is responsible for H4K16ac on all chromosomes in male and female flies

In this study we first wanted to evaluate whether or not MOF is responsible for global H4K16 acetylation, since a recent report focusing on a small set of autosomal gene promoters did not find a decrease in H4K16ac upon disruption of the MOF gene. However, the authors were unable to show that these regions were actual sites of MOF binding [230]. Therefore, in order to provide direct evidence that MOF acts as a HAT also outside the context of the male X chromosome, we performed immunostainings of male and female 3<sup>rd</sup> instar larva polytene chromosomes from wildtype (wt) and *mof2* mutant flies that carry a premature stop codon and lack a functional MOF protein [185], using antibodies against MOF, MSL1 and H4K16ac. While the male X chromosome, marked by MSL1 staining, appeared enriched for H4K16ac in wt flies, we also detected widespread acetylation across autosomes as well as the chromocenter (Figure 4A). Likewise, female samples showed widespread H4K16ac, but no apparent enrichment on the X chromosomes, which also lacked MSL1 staining. Strikingly, H4K16ac was entirely lost from all chromosomes in male and female *mof2* flies, in the absence of MOF. In the males, this was accompanied by reduced staining of MSL1 on the X chromosome in a pattern most likely corresponding to the previously described high affinity sites (HAS), and subsequent delocalization of MSL1 staining to autosomal sites. Importantly, H4K16ac was restored on all chromosomes in both sexes upon expression of an HA-tagged MOF full length transgene (FL-MOF) in flies of the *mof2* background. We recapitulated the same global MOF dependency of H4K16ac by western blots using extracts prepared from 3<sup>rd</sup> instar larvae (Figure 4B). Female wt flies showed about half the amount of global H4K16ac compared to wt males, reflecting the presence of the hyper-acetylated male X chromosome. In the absence of endogenous MOF in the *mof2* background, H4K16ac levels were drastically reduced in males and females, while no significant differences were observed upon probing with antibodies against H4. Importantly, H4K16ac could again be globally restored in both sexes by expression of FL-MOF. This data clearly demonstrates that MOF is responsible for genomewide H4K16ac in *Drosophila*.



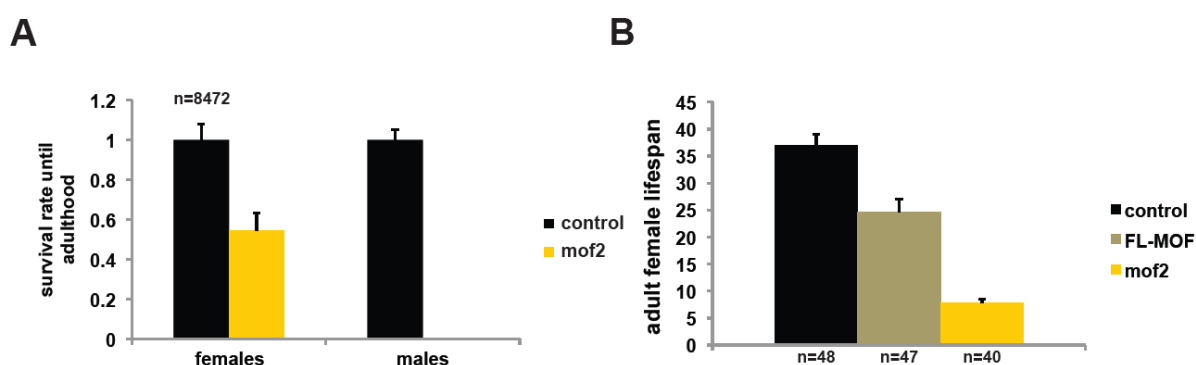
**Figure 4. MOF is the major H4K16 specific HAT in the male and female genome.**

(A) Immunostaining of polytene chromosomes from male and female third instar larva salivary glands, using antibodies against H4K16ac, MSL1 and HA (MOF) as indicated. DNA staining is shown in blue (Hoechst 322). (B) Western blot analysis of extracts prepared from wt, *mof2* and FL-MOF expressing male and female 3<sup>rd</sup> instar larva showing the effect of MOF depletion on H4K16ac

### 3.2 MOF is essential for female survival

Since MOF has been identified in a screen for male specific lethality, the obvious question arises, what the significance of MOF mediated H4K16ac is for female flies. Although a developmental delay and reduced fertility have been described for females homozygous for

the *mof2* allele, that study did not assay for defects in female viability [230]. To address this question we generated homozygous *mof2* females. When compared to control flies, the number of *mof2* females reaching adulthood was reduced by approximately two-fold, and these flies were mostly sterile (Figure 5A). More strikingly however, the average lifespan of *mof2* mutant female flies was drastically reduced to an average of 8 days as compared to 37 days in the control flies with the same set of balancer chromosomes (Figure 5B). Furthermore, this reduction in lifespan was rescued efficiently by expression of a MOF transgene. This result strongly suggests that MOF mediated H4K16ac is essential for female survival, although one cannot rule out additional MOF functions that might contribute to this phenotype.



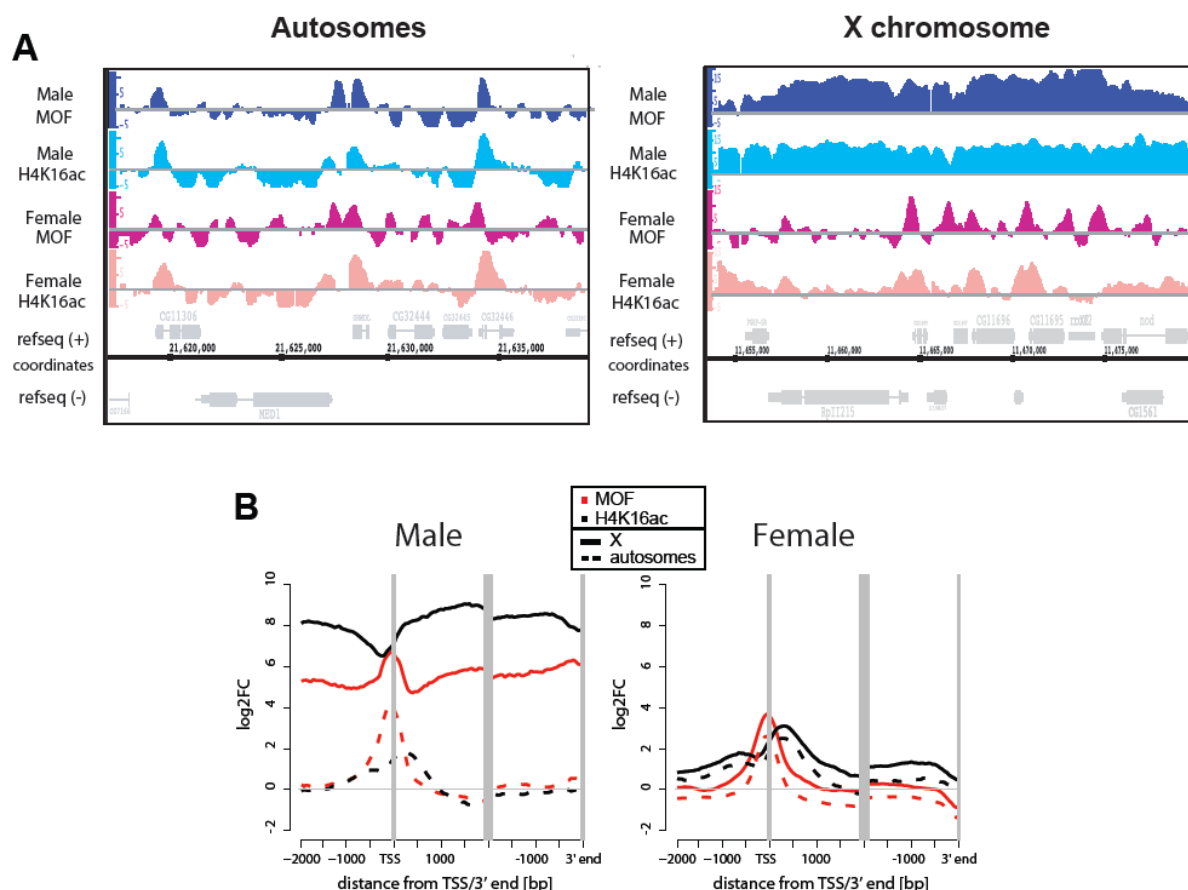
**Figure 5. MOF is essential for female viability.**

(A) Number of eclosed *mof2* adults compared to the control carrying the wt MOF allele. While males show full lethality female number is reduced approx. two fold. Error bars represent the standard deviation from three independent experiments. (B) Lifespan of control, *mof2*, and FL-MOF expressing female flies after eclosion. Error bars represent the standard error of the mean.

### 3.3 Most active genes are targeted by MOF and H4K16ac in both sexes

A previous microarray based study performed in our laboratory correlating genome wide MOF binding and H4K16ac in male and female SL-2 and Kc cell lines was limited to the topmost 1% of signals in order to score the most stringent target sites, and was thus likely to underestimate the real number of MOF targets [215]. We therefore used antibodies against MOF and H4K16ac in chromatin immunoprecipitations followed by ChIP-Seq to obtain comprehensive genome-wide profiles from male and female 3<sup>rd</sup> instar larva salivary glands (Table A1). To control for differences in sonication efficiency and histone occupancy, we also sequenced ChIP material obtained with antibodies against histone H4, as well as the

corresponding input samples. The resulting profiles showed a good overlap between sites of MOF binding and H4K16ac in both sexes (Figure 6A).

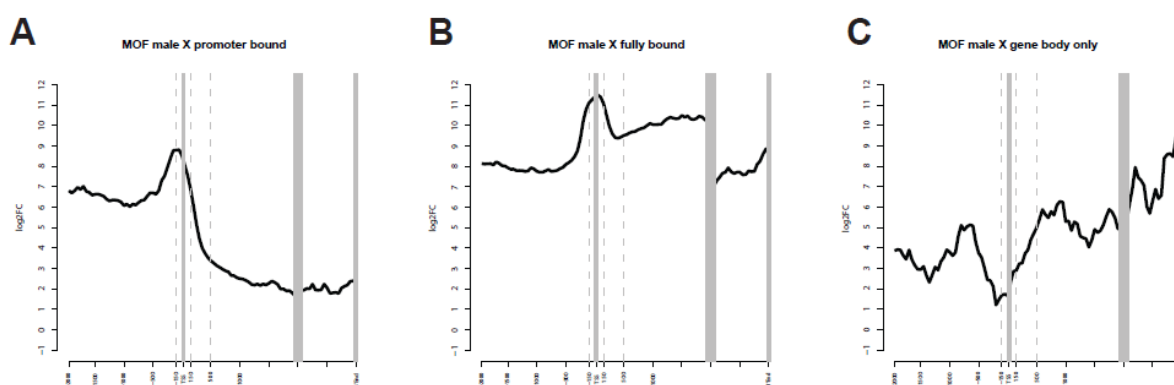


**Figure 6. MOF acetylates promoters across the male and female genome.**

(A) Genomic patterns of MOF-binding and H4K16 acetylation. Screenshots of the Integrated Genome Browser show the patterns (log2FC values) of MOF-binding (dark colour) and H4K16 acetylation (light colour) in male (blue colour shades) and female (red colour shades) salivary glands along a 20kb section of chromosome 3L (left) and a 25 kb section of the X chromosome (right). Genomic coordinates and the locations of annotated Refseq genes on the forward (+) and reverse (-) strands are indicated below. (B) Profiles of MOF-binding and H4K16 acetylation along genes. The plots show the average log2FC signal for MOF-binding (red) and H4K16ac (black) along genes in wt males (left) and females (right). Log2FC signals are shown for the region surrounding the TSS (-2000bp to +2000bp) and for the 3'-end of the gene (3'-end to -2000bp upstream of the transcription termination site). Average log2FC values for X-linked and autosomal genes are represented by solid and dashed lines, respectively.

While the male X chromosome displayed the previously described global enrichment for MOF and H4K16ac [215, 230], MOF and H4K16ac signals on autosomes and the female X chromosome appeared mostly restricted to gene promoters. To confirm that these

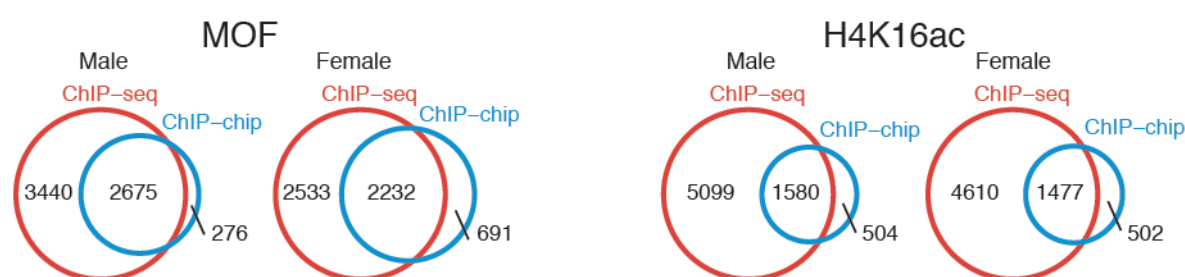
observations apply generally, we computed composite MOF and H4K16ac profiles by averaging the log2 fold-change values across all genes (Figure 6B). These distributions highlight distinct sex and chromosome specific patterns for MOF binding and H4K16ac. On the male X chromosome, MOF binds to the entire gene body, with peaks of increased binding at the promoter and the 3'-end of genes. On male autosomes and generally in females, MOF binds only at gene promoters. The pattern of H4K16 acetylation closely follows that of MOF binding [215]. Interestingly, we observe a shift of the H4K16ac signal downstream of gene promoters, most likely reflecting acetylation of the first nucleosome downstream of the transcription start site (TSS). We next used the pattern of MOF-binding to classify genes into three categories: (i) those with no binding; (ii) those that are bound in the promoter region; (iii) those that are fully bound along their whole length (table A2, for details see materials and methods). On the male X chromosome, 898 genes were fully bound by MOF, while 1,704 genes were fully acetylated, corresponding to 75% and 94% of all active genes, respectively. In contrast, only one X-linked gene was fully bound by MOF in females and just 158 were fully acetylated, demonstrating that MOF binding and acetylation in the transcribed region of X-linked genes are specific to dosage compensated genes in males. We also detected 387 genes on the male X chromosome that were bound by MOF only at their promoters and a much smaller group of 44 genes that showed MOF binding only in their transcribed regions (Figure 7).



**Figure 7. Profiles of MOF binding along X chromosomal genes**

Average log2FC signal for MOF-binding along X chromosomal genes in wt males classified as being bound only at gene promoters (A), gene promoters and full bodies (B) or exclusively at gene bodies (C). Log2FC signals are shown for the region surrounding the TSS (-2000bp to +2000bp) and for the 3'-end of the gene (3'-end to -2000bp upstream of the transcription termination site).

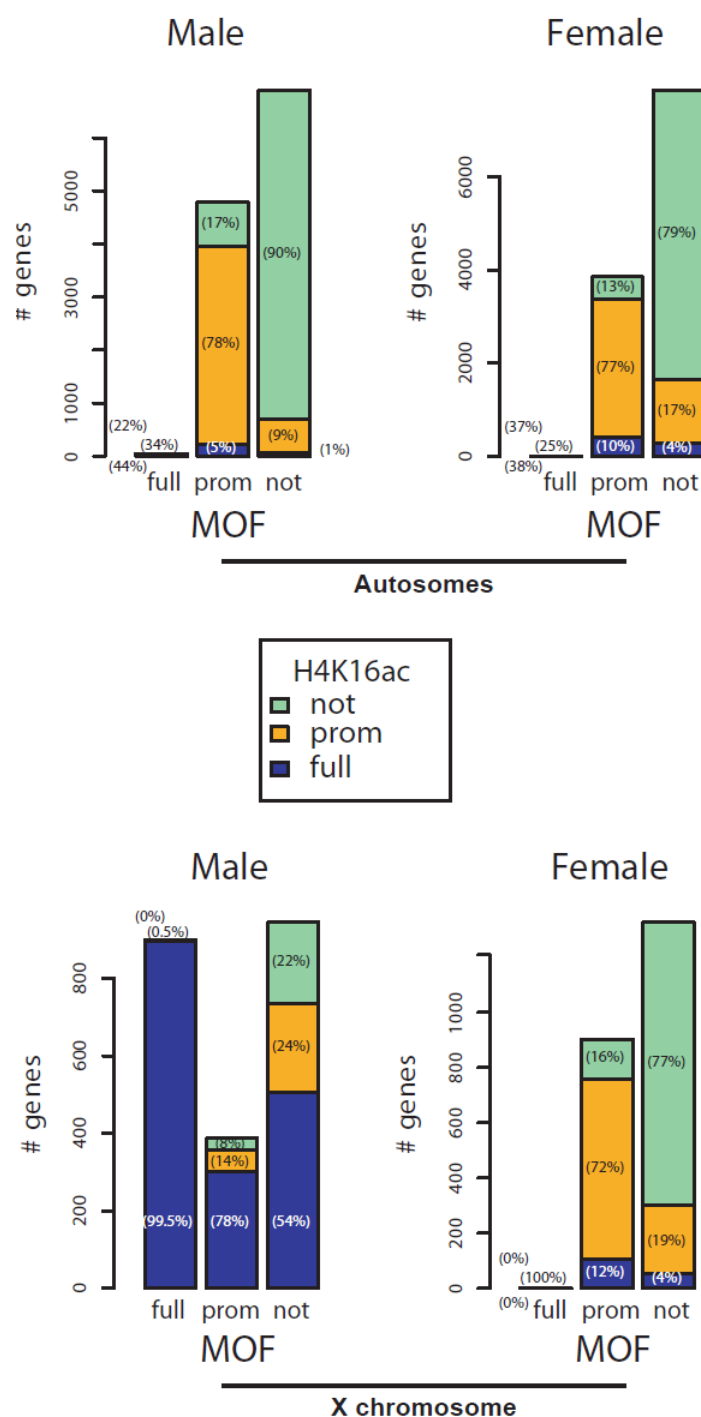
Interestingly, most gene promoters of this latter group also appeared devoid of NSL proteins in our earlier analysis [227], again demonstrating that MOF binding to gene promoters happens in the context of the NSL complex (data not shown). Importantly, we also detected a large number of MOF bound (4830) and acetylated (4690) autosomal targets, with most of the signals located at gene promoters. This number includes a striking 81% (MOF) and 76% (H4K16ac) of all active autosomal genes, signifying the role of MOF for autosomal gene regulation. Similar numbers were found across all chromosomes in females. Genes positive for MOF binding and H4K16ac included most targets previously identified by Kind et al., however, due to the higher sensitivity of the ChIP-Seq technique the overall number of positive genes was more than doubled in either case (Figure 8).



**Figure 8. The ChIP-Seq approach is more sensitive compared to ChIP-chip**

Overlap of MOF-bound and H4K16-acetylated genes detected by the ChIP-chip analyses performed by Kind et al. (2008) from cell lines and our ChIP-seq analysis using third instar salivary gland samples. Both male and female samples are presented.

Further evidence that H4K16ac at gene promoters is indeed mediated by MOF was provided by the good correlation between MOF binding and H4K16ac at gene promoters in both sexes (Figure 9). On autosomes, 83% (males) and 87% (females) of MOF-bound genes were also acetylated. Similarly, on the X chromosome, 98% of the male and 85% of the female MOF-bound genes were acetylated at the same time. However, on the male X chromosome, we also detected 735 genes that were acetylated despite lacking MOF, suggesting spreading of MOF mediated H4K16 acetylation into neighboring genes. This is in agreement with a model in which transient DCC interactions can spread H4K16ac in large domains across the male X chromosome, even in the absence of detectable DCC binding [196, 215, 230]. We did not observe the same phenomenon on the female X chromosome.

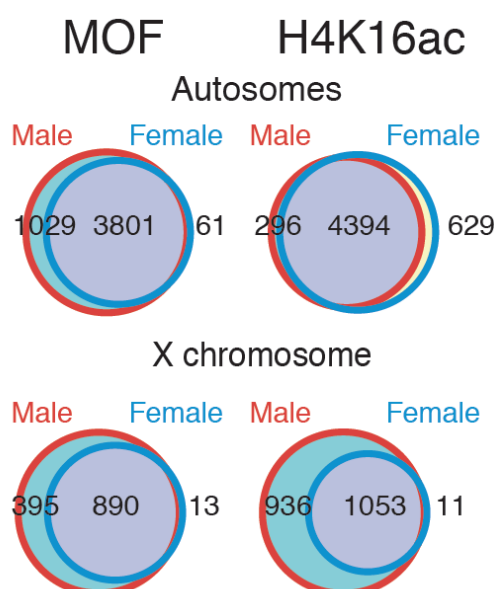


**Figure 9. MOF correlates with H4K16ac genome wide.**

Overlap in MOF-bound (left) and H4K16-acetylated (right) autosomal (top) and X-linked genes (bottom) genes between males and females. (D) Overlap of MOF-bound with H4K16-acetylated genes in wt males and females. Autosomal (left) and X-linked (right) genes were divided into three categories showing full (full), promoter-only (prom) or no (not) binding by MOF. For each category, the stacked barplot specifies the number of genes showing full (full), promoter-only (prom) or no (not) H4K16 acetylation. The numbers within each block indicate the percentage of genes within this block relative to all genes in the category.



Next, we compared MOF binding and H4K16ac between males and females. At X-linked promoters, we found a high level of correlation between MOF binding in males and females ( $\rho=0.72$ ). However, this correlation was lost when binding was compared for gene bodies ( $\rho=0.05$ ), where MOF binding is specific for males. Similarly, H4K16ac correlated at gene promoters ( $\rho=0.54$ ), but not along the body of genes ( $\rho=0.27$ ). When comparing individual genes, we found that almost all female MOF-bound genes on the X chromosome (99%) are also MOF targets in males (Figure 10). A similar result was observed for H4K16ac. However, a significant proportion of genes was MOF-bound or acetylated (31% and 47%, respectively) only in males, reflecting male specific action of the DCC. On the autosomes, correlations between the sexes were high at promoters for both MOF and H4K16ac ( $\rho=0.77$  and  $\rho=0.86$  respectively). Additionally, we found large overlaps between individual target genes: 98% of all autosomal MOF-bound genes in females were also bound by MOF in males, and 87% of the acetylated genes in females were also acetylated in males. These results suggest that outside the context of the DCC, MOF functions very similar in male and female flies.

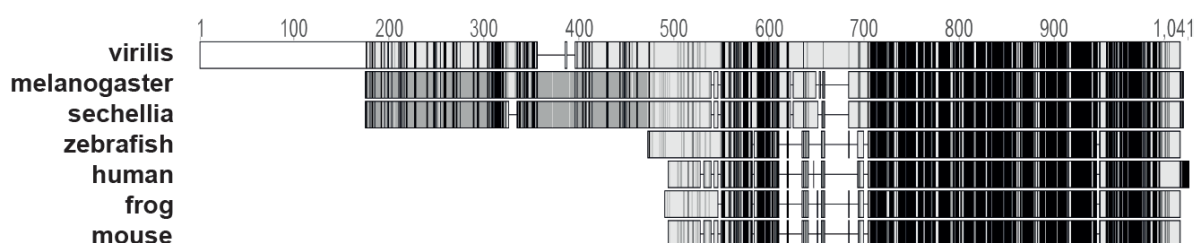


**Figure 10. MOF regulates the same genes in male and female flies.**

Overlap in MOF-bound (left) and H4K16-acetylated (right) autosomal (top) and Xlinked genes (bottom) genes between males and females.

### 3.4 The MOF N-terminus controls DCC function

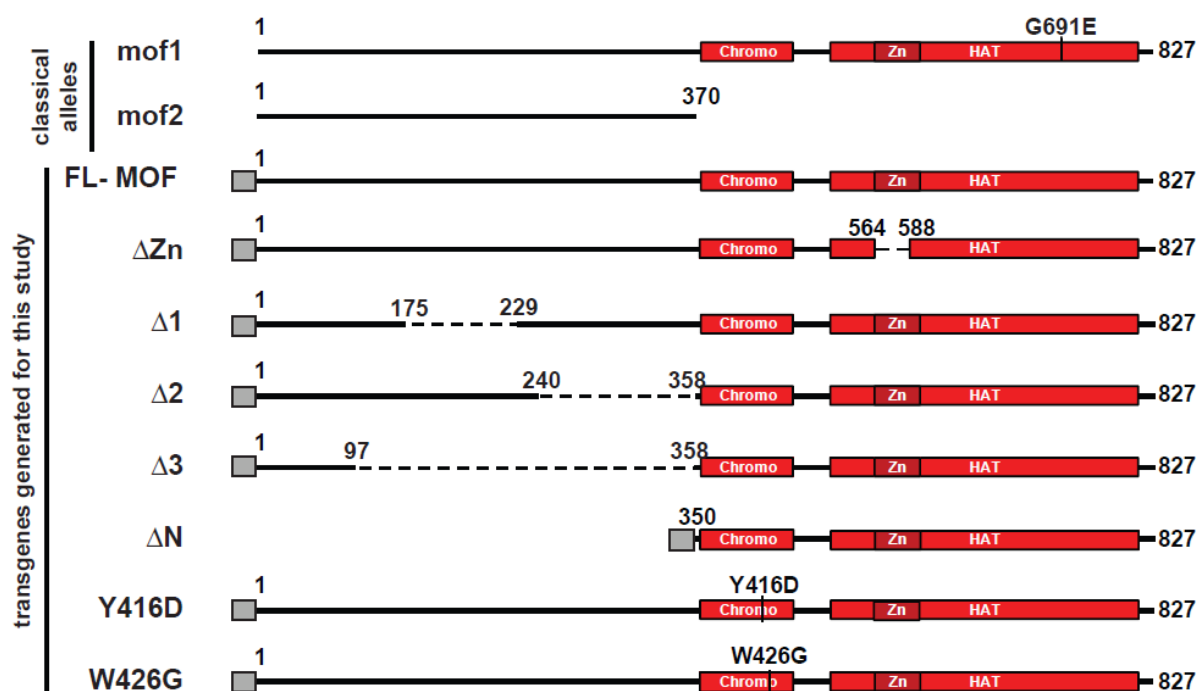
The above analysis underscored the dual role that MOF plays in transcription regulation and dosage compensation in *Drosophila*. The increased diversity of MOF functions in *Drosophila* species implies the necessity for tight regulation of MOF targeting and activity. We therefore wanted to dissect the molecular mechanisms that orchestrate these functions to utilize this evolutionary conserved transcriptional regulator for compensation of the male X chromosome. A hallmark of MOF in *Drosophila* species is the presence of a long N-terminal segment that is absent in other organisms (Figure 11).



**Figure 11. The MOF N-terminus is specific to *Drosophila* species**

Alignment of the MOF amino acid sequence in different animal species. Evolutionary distance between *Drosophila melanogaster* and *D. virilis* is approx. 55 mio years, *D. melanogaster* and *D. sechellia* approx. 2,5 mio years. Sequence conservation is indicated by darker greyscales.

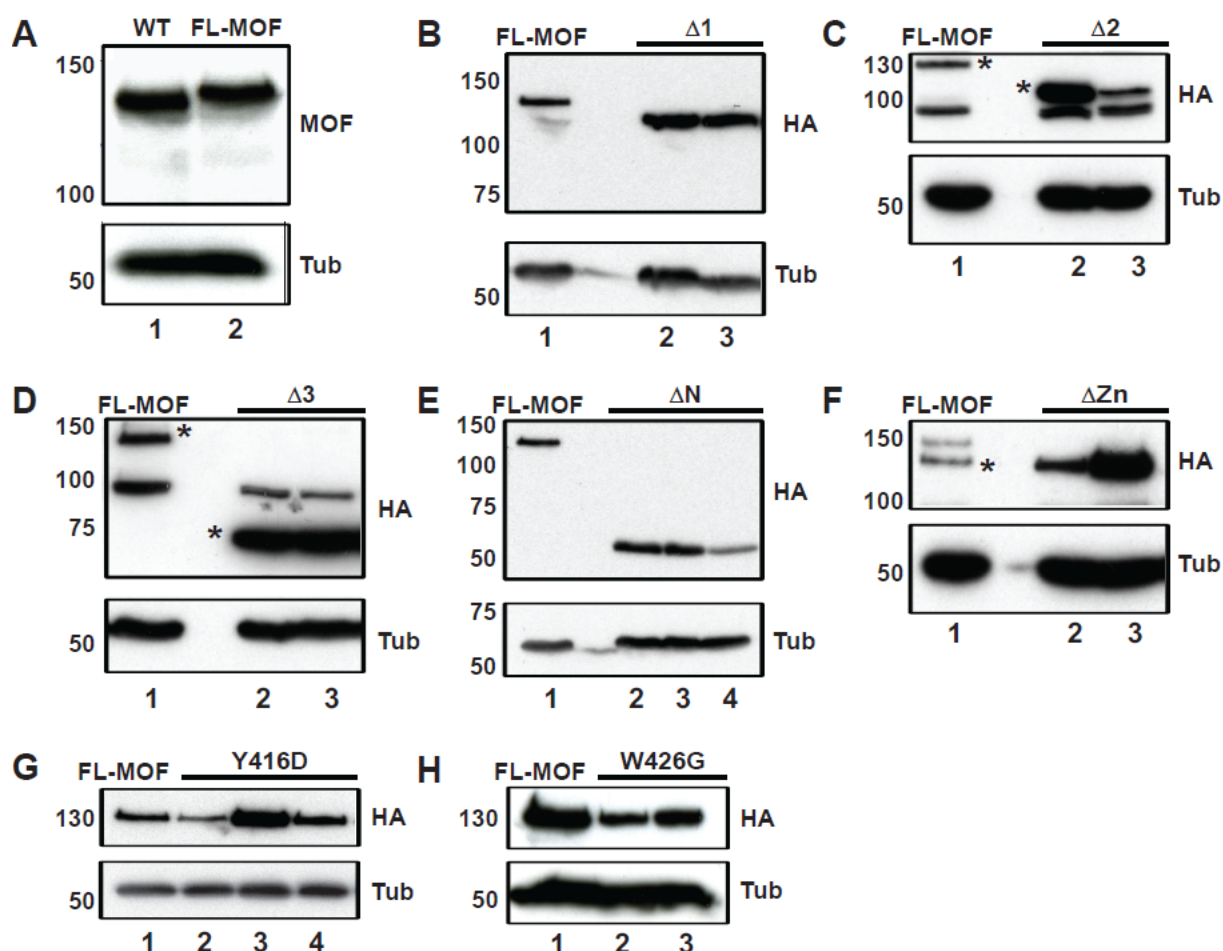
This region shows relatively little sequence conservation and is variable in length between different *Drosophila* species. In *Drosophila melanogaster* the first 350 amino acids of MOF do not contain any known protein domains and appear as largely unstructured in secondary structure predictions. So far no function has been assigned to this region, which comprises about half of the MOF protein; however the absence of this part in organisms with different modes of dosage compensation suggest a specific requirement of the MOF N-terminus for this process. To test if the N-terminus carries essential functions *in vivo*, we assayed the capability of a series of MOF mutant proteins to rescue the male lethal phenotype associated with the loss of endogenous MOF.



**Figure 12. MOF transgenes analyzed in this study**

(A) Schematic representation of the domain structure of the known MOF alleles *mof1* and *mof2*, as well as the MOF derivatives generated for this study. Globular domains in the MOF protein are the chromobarrel domain (chromo), the C-terminal HAT domain (HAT) containing a zinc finger (Zn). G691 is located in the catalytic center of the MOF HAT domain. Y416D and W426G reside in the chromobarrel domain and disrupt RNA binding.

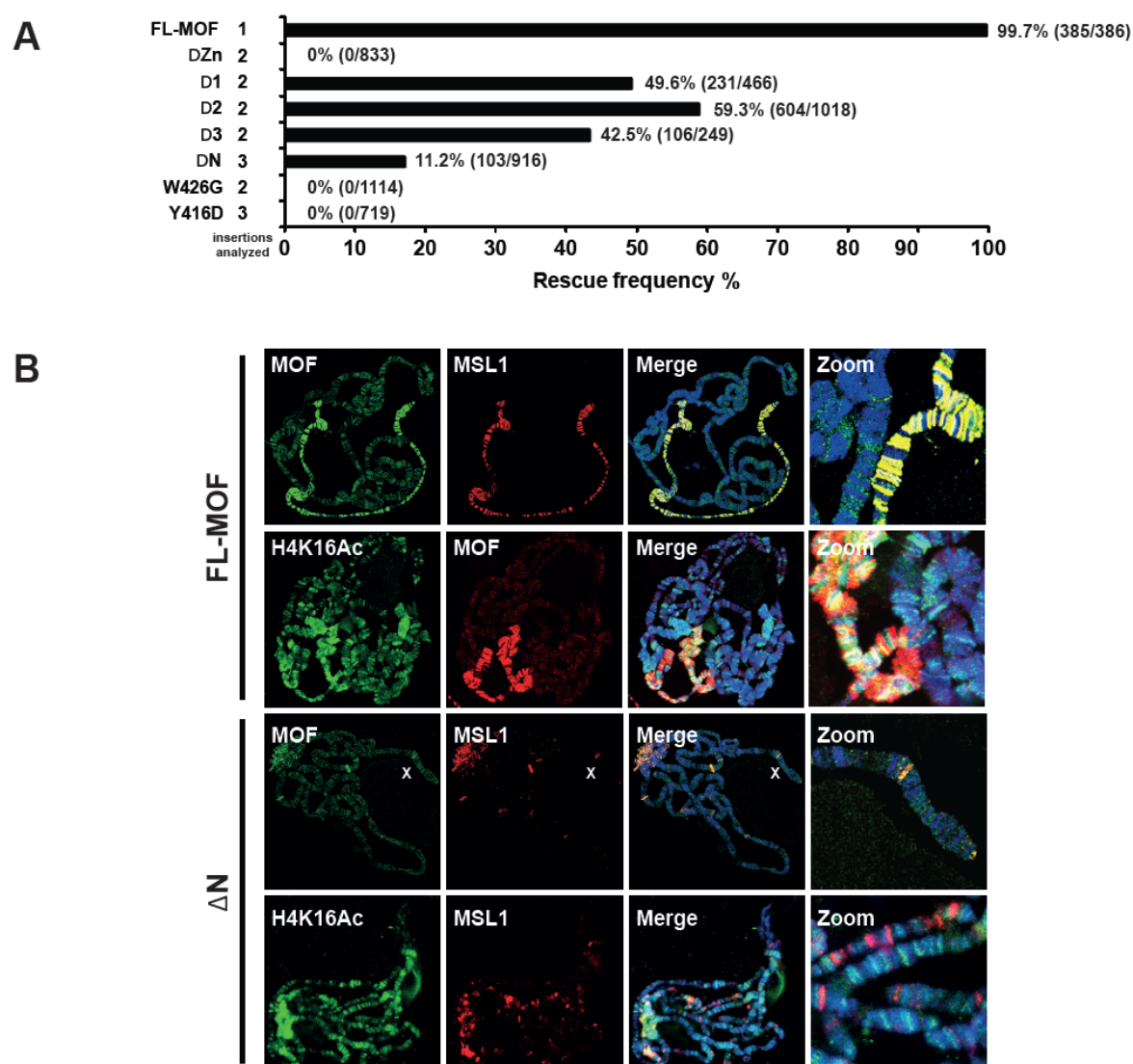
To this end we introduced HA-tagged MOF transgenes by p-element mediated transformation into flies, including a control construct comprising full length MOF (FL-MOF) and a series of deletions in the N-terminal half of the MOF protein,  $\Delta$ 1 ( $\Delta$  176-228),  $\Delta$  2 ( $\Delta$  241-357),  $\Delta$  3 ( $\Delta$  98-357) and  $\Delta$ N ( $\Delta$  1-349) (Figure 12). All transgenes were expressed to wt levels upon induction with armadillo-Gal4, as determined by western blot (Figure 13). Full length HA-tagged MOF completely rescued male lethality in this assay when compared to heterozygous *mof2* females resulting from the same cross (Figure 14A). In contrast, male viability was increasingly compromised up to ~90% upon progressive deletions of the N-terminal region. This result demonstrates that this part of the protein is essential for MOF function. Having confirmed the essential nature of the N-terminal domain for MOF function, we wanted to assay for defects in MOF targeting to X-linked genes or autosomes, DCC recruitment to the X chromosome, and general H4K16ac.



**Figure 13. MOF transgenes are expressed to wt levels.**

(A-H) Western blot analysis using extracts from male adult fly heads expressing the indicated transgenes. Two to three independent lines were analyzed for each mutant variant. Tubulin (Tub) was used as a loading control.

We again performed immunostainings of 3<sup>rd</sup> instar larva polytene chromosomes from male flies that express MOF transgenes in the *mof2* background. Upon immunostaining with anti-HA antibodies, FL-MOF appeared in a wildtype pattern, showing pronounced enrichment on the X chromosome but also clear targeting to all autosomes (Figure 14B). This was accompanied by widespread H4K16ac, which appeared enriched on the X chromosome, while MSL1 staining remained restricted to the X, all reflecting wt DCC targeting and function. This data once more confirmed the functionality of the full length HA-tagged MOF protein *in vivo*. Strikingly, in the absence of the MOF N-terminus, preferential targeting of MOF to the male X chromosome was entirely lost. Also the staining of autosomal bands appeared reduced. Furthermore, we detected ectopic binding of MOF to the chromocenter.

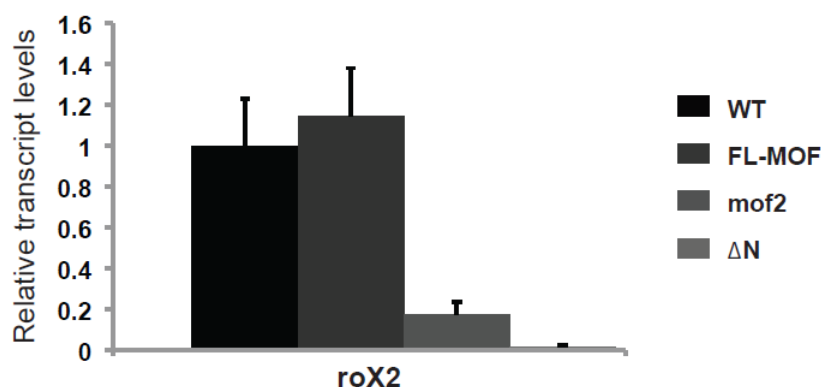


**Figure 14. The N-terminal domain of MOF controls DCC function**

(A) Rescue of male viability by expression of MOF transgenes in the *mof2* background. Percentages refer to the number of *mof2* males compared to the number of heterozygous *mof2* females from the same cross. (B) Immunostaining of polytene chromosomes from *mof2* male third instar larvae salivary glands expressing FL-MOF and  $\Delta N$  MOF transgenes. Antibodies against HA (MOF), MSL1 and H4K16ac were used as indicated in the figure. DNA staining is shown in blue (Hoechst 322).

At the same time MSL1 staining was reminiscent of the patterns that have been observed in the absence of both roX RNAs [191], with a nearly complete delocalization of MSLs from the X chromosome, accompanied by severe ectopic binding to autosomal sites and the chromocenter. Indeed, when we measured the levels of roX RNAs in the  $\Delta N$  mutant background we found a severe depletion of roX2 by more than 98%, suggesting that the N-terminus of MOF is required for proper incorporation of roXs into the DCC (Figure 15). This reduction was even more severe than the one observed in the absence of MOF, where core

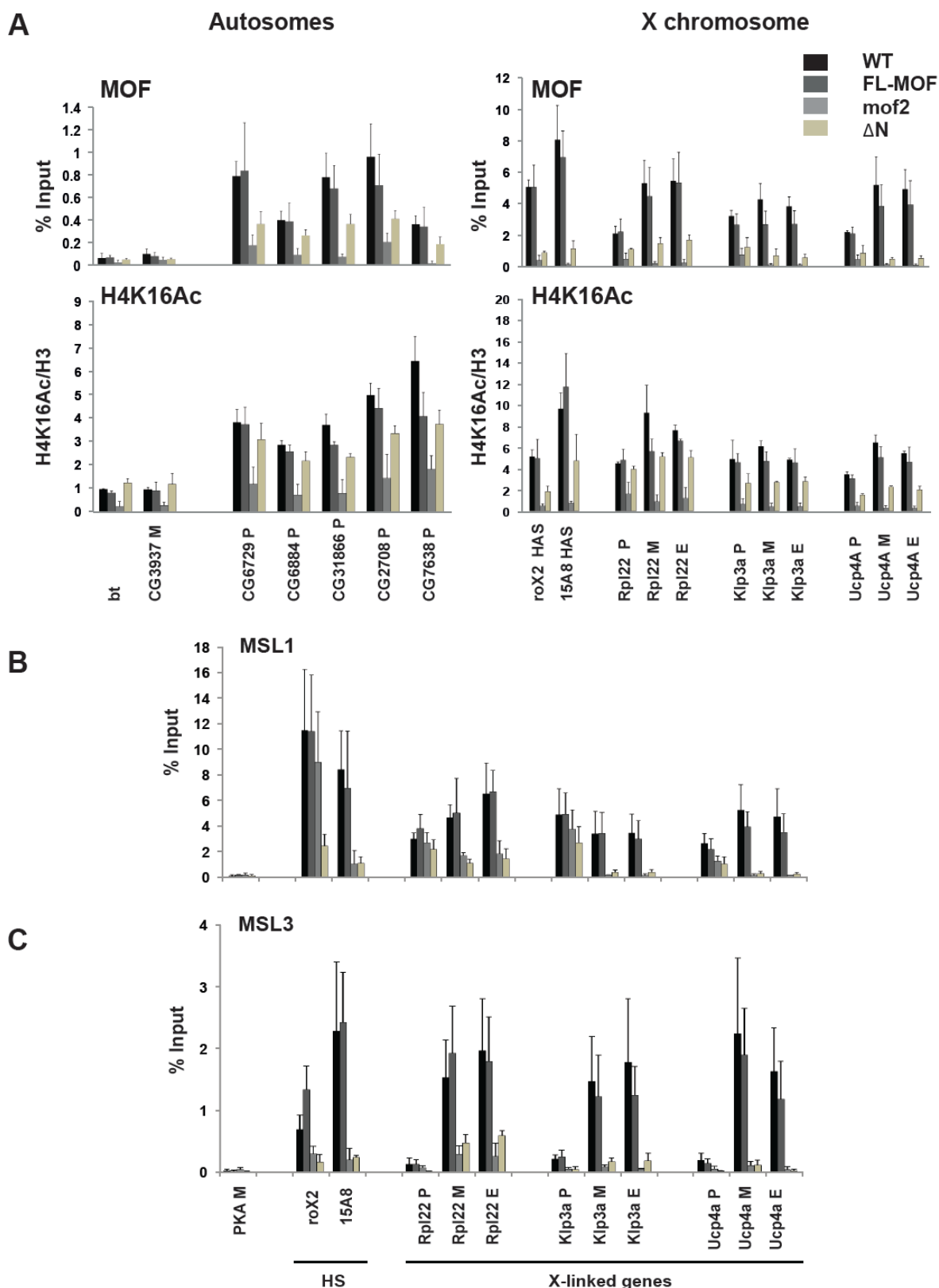
MSL complexes reside together with roX RNAs at high affinity sites, suggesting that roX RNAs cannot be incorporated into the DCC when a truncated MOF protein is present in the complex.



**Figure 15. Degradation of roX RNA in the absence of the MOF N-terminus.**

roX2 levels in *mof2* male larva salivary glands expressing MOF transgenes as indicated. Transcript levels were normalized to roX2 DNA recovered from the same sample.

Surprisingly however, although general chromatin targeting of  $\Delta N$  MOF appeared reduced, we still detected widespread H4K16ac on all chromosomes (Figure 14B bottom panel). To get a more detailed insight into the defects in autosomal vs. X chromosome targeting of the  $\Delta N$  MOF protein, we performed ChIP from male 3<sup>rd</sup> instar larva to monitor MOF binding, H4K16ac and DCC recruitment at higher resolution. Chromatin was immunoprecipitated, using antibodies against MOF, H4K16ac, MSL1 and MSL3. The recovered DNA was measured by quantitative real-time PCR (qPCR). We assayed MOF binding and H4K16ac at the promoter regions of the autosomal genes *cg6729*, *cg6884*, *cg31866*, *cg2708* and *cg7638* and at two non-targets, *bt* and *cg3937*. To monitor MOF function on the X chromosome, we used the known high affinity sites (HAS) at the roX2 gene and an additional HAS at the cytological location 15A8, which had previously been identified in an MSL3 mutant background [199]. We also included sites at the promoter, middle and 3'end of three X-linked genes, *Rpl22*, *Klp3a* and *Ucp4a*. The first one of these, *Rpl22*, previously showed some MSL binding in the absence of MSL3 and can thus be described as a medium affinity site, while the remaining two genes are low affinity MSL targets [199].



**Figure 16. The MOF N-terminus is required for MSL targeting**

(A) ChIP using MOF (upper panel), and H4K16ac (lower panel) antibodies in wt male 3<sup>rd</sup> instar larva as well as *mof2* larva or *mof2* larva that express FL-MOF or  $\Delta$ N MOF transgenes. Binding to the autosomal genes *bt*, *cg3937*, *cg6729*, *cg6884*, *cg31866*, *cg2709*, *cg7638*, the X-linked high affinity

---

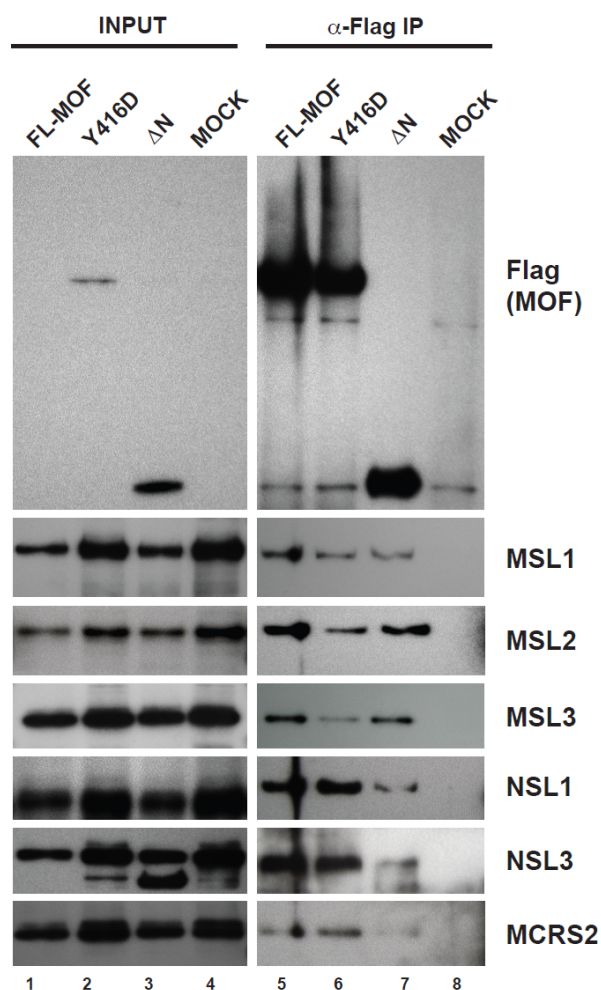
sites at the *roX2* gene and location 15A8, as well as the X chromosomal genes *Rpl22*, *Klp3a*, and *Ucp4a* is shown. Primers were positioned at the promoter (P), middle (M), and end (E) of genes. The exact position of the primers is described in the Supplemental Data. ChIP is shown as percentage recovery of input DNA (% Input). Error bars represent standard deviation (StDev) of three independent experiments. (B) FL-MOF fully restores MSL binding. ChIP using MSL1 antibodies in wt male 3<sup>rd</sup> instar larva as well as *mof2* larva or *mof2* larva that express FL-MOF or  $\Delta$ N MOF transgenes. Binding to the X-linked high affinity sites at the *roX2* gene and location 15A8, as well as the X chromosomal genes *Rpl22*, *Klp3a*, and *Ucp4a* is shown. PKA is used as a negative control. Primers were positioned at the promoter (P), middle (M), and end (E) of genes. (C) ChIP as in (B) using antibodies against MSL3.

---

The pattern of FL-MOF binding in the *mof2* background was indistinguishable from the one observed for endogenous MOF in the wt, with clear binding to autosomal promoters, to the promoter and transcribed region of X-linked genes and to HAS (Figure 16A). MOF binding was accompanied by H4K16ac at all of these loci in the wt and FL-MOF background. Upon removal of endogenous MOF in the *mof2* background alone, MOF binding and H4K16ac were both lost from all sites tested. Accordingly, targeting of MSL1 and MSL3 was lost from the body of X-linked genes in the *mof2* background, but was restored to the wt pattern upon expression of FL-MOF (Figure 16B, C). Interestingly, the 15A8 HAS, which had been identified in an MSL3 mutant background, was no longer bound by MSL1 in the absence of MOF, demonstrating the qualitative differences among high-affinity sites in varying genetic backgrounds [232]. Confirming immunostainings, binding of the truncated  $\Delta$ N MOF was lost from the transcribed regions of low affinity target genes on the X chromosome, but also strongly diminished at X-linked promoters and HAS (Figure 16A). Compromised MOF binding was accompanied by a loss of MSL1 and MSL3 from the same target sites (Figure 16B, C). Importantly, MSL1 was also lost from the *roX2* high affinity site, most likely as a result of low *roX* RNA levels and compromised DCC assembly. Chromatin binding of  $\Delta$ N MOF was also approximately twofold reduced at the promoters of X-linked and autosomal genes. We therefore asked if reduced promoter binding of  $\Delta$ N MOF reflects compromised integration into the NSL complex. To address this question, we generated SL-2 cell lines stably expressing Flag-tagged FL MOF, and  $\Delta$ N MOF transgenes. After immunoprecipitation with anti-Flag antibodies followed by western blotting, membranes were probed against Flag, NSL1, NSL3 and MCRS2, as well as MSL1, MSL2 and MSL3. In this assay, FL MOF was stably interacting with all other complex partners. Deletion of the MOF N-terminus did not have an apparent effect on the interactions with individual MSL proteins. Importantly



however, interactions with all three members of the NSL complex were clearly diminished, consistent with reduced targeting of  $\Delta N$  MOF to gene promoters (Figure 17). Together, these results suggest a dual function of the *Drosophila* specific N-terminus of MOF, which controls MOF integration into the NSL complex as well as DCC assembly on the male X chromosome.

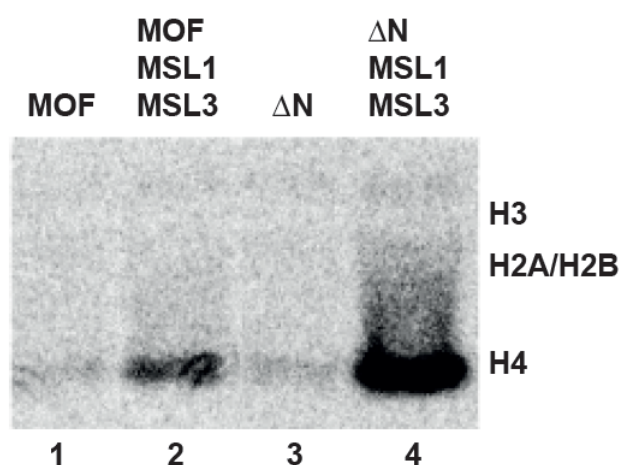


**Figure 17. The MOF N-terminus is required for incorporation into the NSL complex, provided by Erinc Hallacli.**

Coimmunoprecipitations were performed from nuclear extracts obtained from SL-2 cell lines stably expressing FLAG-tagged wt, Y416D and  $\Delta N$  MOF. Input material and immunocomplexes were subjected to western blot analysis using the indicated antibodies for detection.

### 3.5 The N-terminus constrains MOFs enzymatic activity

The high levels of H4K16ac that we detected in the presence of  $\Delta$ N MOF *in vivo*, although chromatin targeting of the enzyme itself was reduced, suggested a potentially altered activity level of  $\Delta$ N MOF. We therefore wanted to directly measure the enzymatic activity of the  $\Delta$ N MOF protein. To this end we generated baculoviruses carrying the HA-tagged FL MOF and  $\Delta$ N MOF constructs to induce expression in SF21 cells for subsequent purification of the recombinant proteins. We then performed *in vitro* HAT assays on purified endogenous nucleosomes, using FL MOF and  $\Delta$ N MOF proteins that were copurified with MSL1 and MSL3 to yield enzymatically active trimeric complexes [231]. In this assay,  $\Delta$ N MOF containing trimeric complexes showed an about five-fold increase in acetylation activity compared to FL MOF, while monomeric FL MOF and  $\Delta$ N MOF had the same low levels of residual activity (Figure 18). This result suggests that an autoregulatory function in the N-terminal domain of MOF constrains its enzymatic activity, and explains why levels of H4K16ac remain high *in vivo* although targeting of  $\Delta$ N MOF to chromatin is impaired. Compromised DCC targeting to the X chromosome in the presence of  $\Delta$ N is thus likely the result of defects in complex assembly and roX degradation *in vivo* rather than of a reduction in H4K16ac.

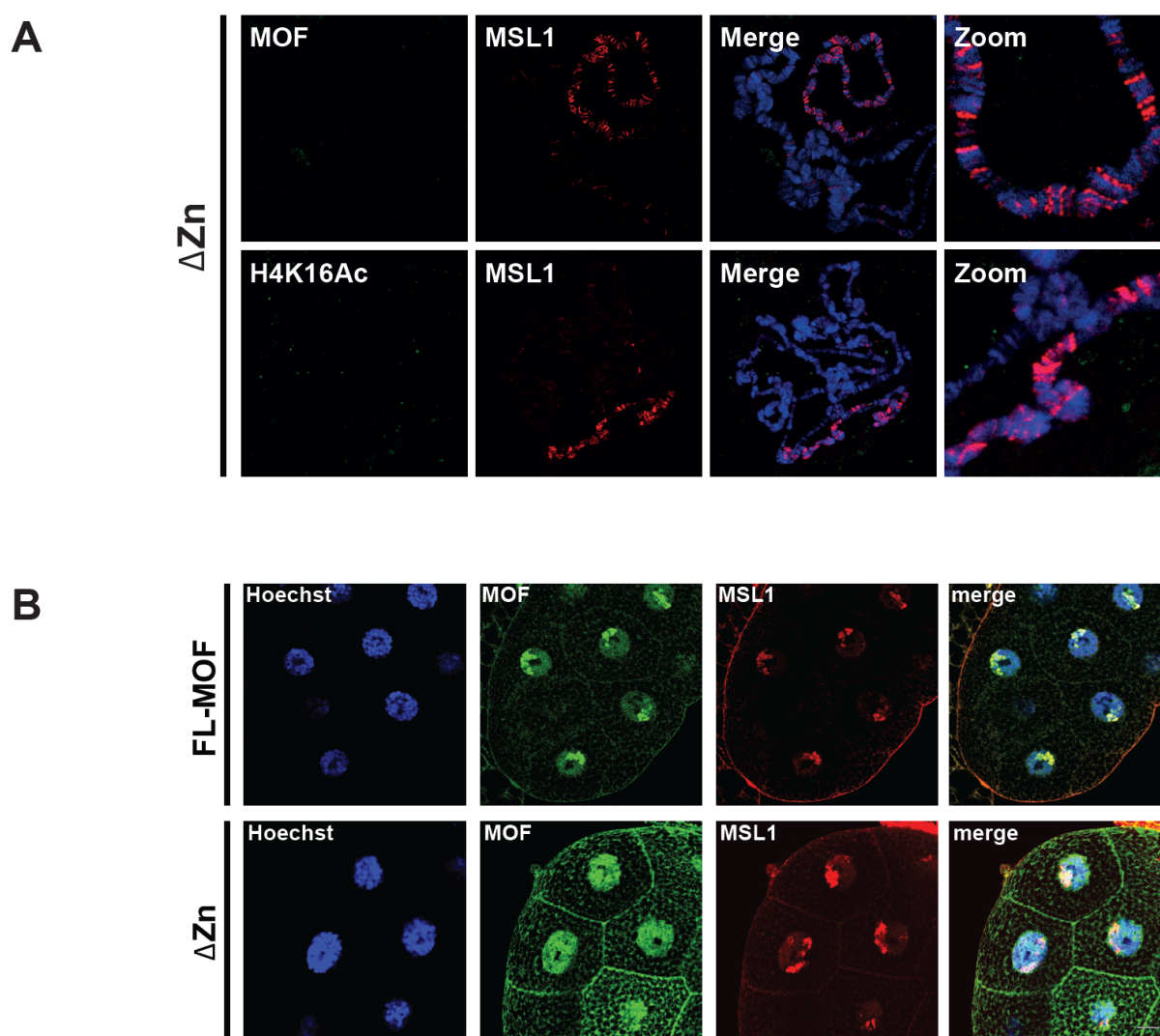


**Figure 18. The MOF N-terminus constrains HAT activity, provided by Herbert Holz.**

Acetylation assay on native nucleosomes. Each reaction contains 5 nM of FL or  $\Delta$ N MOF proteins, with or without MSL1 and MSL3, respectively, and 1.5  $\mu$ g of native nucleosomes purified from MCF-7 cells.

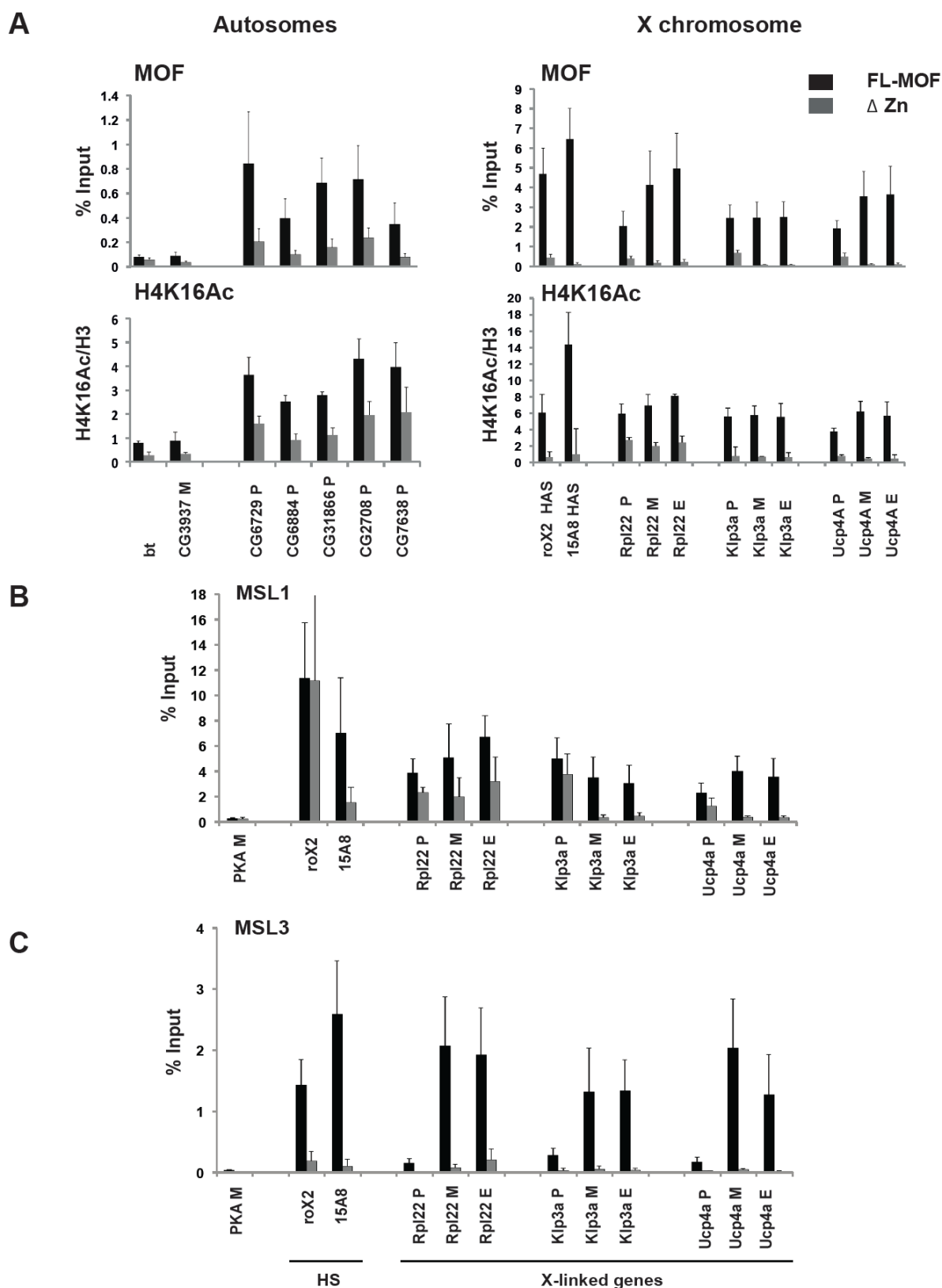
### 3.6 The Zn finger of MOF is required for targeting to the X and autosomes

Previous work had identified the zinc finger region in the HAT domain of MOF as required for interaction with MSL1 and for nucleosome binding [231, 233]. Accordingly, an overexpressed MOF variant with a disrupted zinc finger lost the characteristic enrichment on the X chromosomal territory upon immunostaining in male SL-2 cells. Recently, the same region of MOF has been shown to mediate interactions with the NSL complex via NSL1 [232]. We therefore wanted to ask if the zinc finger cooperates with the N-terminal domain in MOF regulation, and if disruption of the zinc finger would thus lead to similar phenotypes.



**Figure 19. The zinc finger region is required for general chromatin targeting**

(A) Immunostaining of polytene chromosomes from *mof2* male third instar larvae salivary glands expressing the  $\Delta Zn$  MOF transgene, using antibodies against H4K16ac, MSL1 and HA (MOF) as indicated. DNA staining is shown in blue (Hoechst 322). (B) Immunostaining of salivary glands from *mof2* male third instar larvae expressing the FL-MOF or  $\Delta Zn$  MOF transgene, using antibodies against HA (MOF) and MSL1 as indicated. DNA staining is shown in blue (Hoechst 322).



**Figure 20. The zinc finger region is required for chromatin binding of MOF**

(A) Chromatin immunoprecipitation (ChIP) using MOF (upper panel), and H4K16ac (lower panel) antibodies in *mof2* male 3<sup>rd</sup> instar larva expressing FL-MOF (black) or ΔZn MOF transgenes (grey). Binding to the autosomal genes *bt*, *cg3937*, *cg6729*, *cg6884*, *cg31866*, *cg2709*, *cg7638*, the X-linked high affinity sites at the *roX2* gene and location 15A8, as well as the X chromosomal genes *Rpl22*,

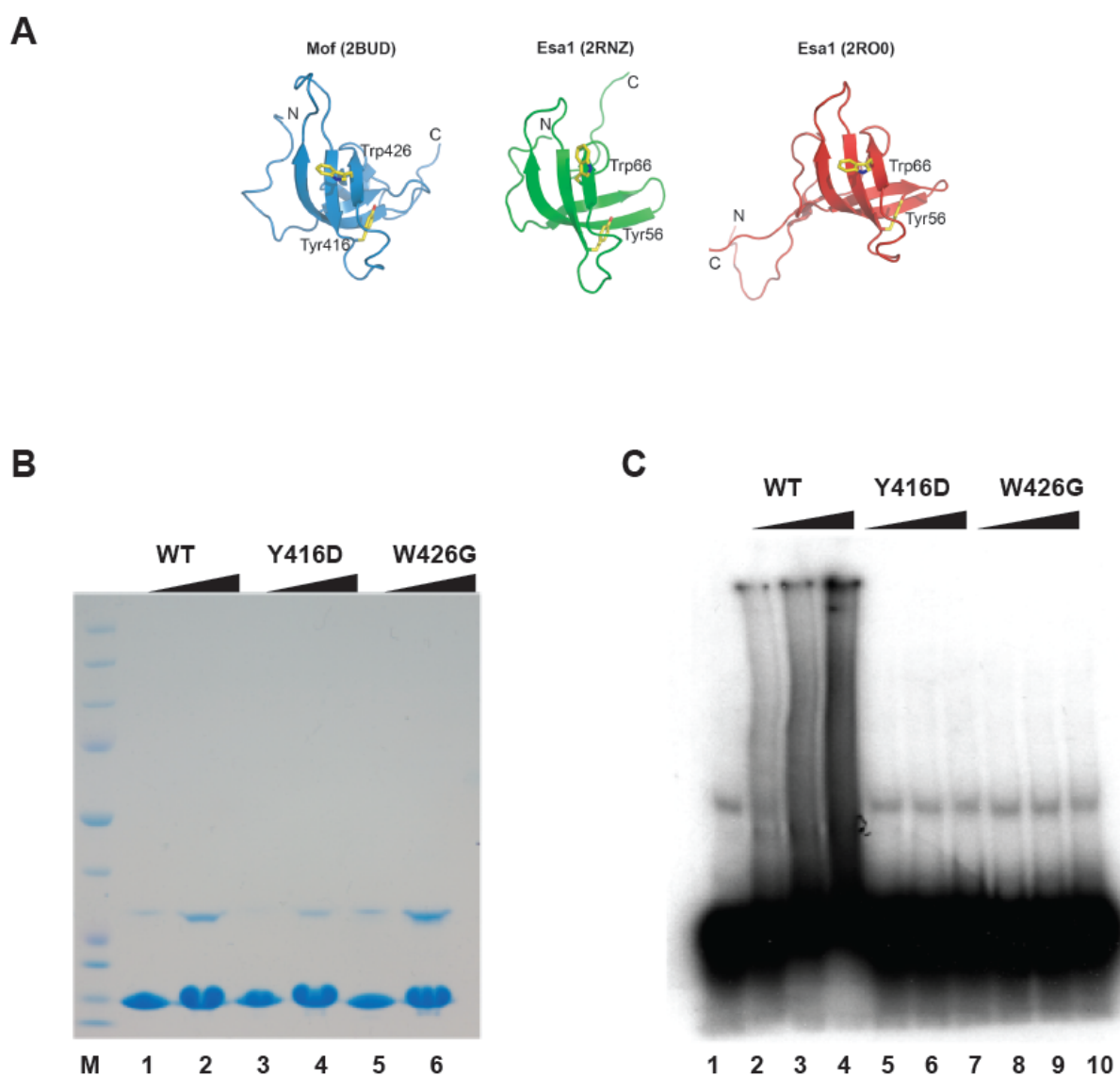
*Klp3a*, and *Ucp4a* is shown. Primers were positioned at the promoter (P), middle (M), and end (E) of genes. The exact position of the primers is described in the Supplemental Data. ChIP is shown as percentage recovery of input DNA (% Input). Error bars represent standard deviation (StDev) of three independent experiments. (B) MSL1 recruitment to the male X chromosome is lost in the presence of  $\Delta$ Zn MOF. ChIP using MSL1 antibodies in *mof2* male 3<sup>rd</sup> instar larva expressing FL-MOF (black) or  $\Delta$ Zn MOF transgenes (grey). Binding to the X-linked high affinity sites at the *roX2* gene and location 15A8, as well as the X chromosomal genes *Rpl22*, *Klp3a*, and *Ucp4a*, as well as the autosomal *PKA* gene is shown. *PKA* is used as a negative control. Primers were positioned at the promoter (P), middle (M), and end (E) of genes. (C) ChIP as in (B) using antibodies against MSL3.

A construct carrying a short deletion comprising the zinc finger region,  $\Delta$ Zn MOF ( $\Delta$ 565-587), was unable to rescue the male specific phenotype of the *mof2* background, confirming the essential nature of this domain (Figure 13 and 14A). However, the phenotypes observed upon immunostaining differed markedly from the ones observed in  $\Delta$ N MOF. Suggesting a general defect in chromatin binding,  $\Delta$ Zn MOF was no longer detectable on polytene chromosomes (Figure 19A), while at the same time immunostaining of whole mount salivary glands verified the nuclear localization of the  $\Delta$ Zn MOF protein (Figure 19B). Consistently, failure of  $\Delta$ Zn MOF targeting led to a complete loss of H4K16ac staining from all chromosomes. As a consequence, MSL1 appeared restricted to HAS on the X chromosome and showed enhanced binding to autosomes, corresponding to the pattern observed in the *mof2* background alone. This result was verified by ChIP, showing a complete loss of  $\Delta$ Zn MOF from all target sites (Figure 20A). Disrupted chromatin binding was accompanied by a concomitant loss of H4K16ac, and consequently compromised spreading of MSL1 and MSL3 into X-linked genes (Figure 20A, B, C). These data suggest that the zinc finger region is required for basic functionality of MOF, with respect to all chromatin related tasks. This is in marked contrast to the N-terminal domain, which rather controls the distribution and activity of MOF to carry out its diverse functions.

### 3.7 The chromobarrel domain is required for global H4K16ac

Two point mutations in the MOF chromobarrel domain at Tyr416 and Trp426 were previously shown to disrupt MOF's interaction with roX RNA *in vitro* and *in vivo* [236], and the corresponding residues have been shown to be required for nucleic acid binding in Esa1, a closely related HAT in yeast (Figure 21A). To verify that the MOF chromobarrel domain directly interacts with RNA, we generated plasmids containing the chromobarrel domain of

MOF (amino acids 346-448) with an N-terminal 6 His-tag for bacterial expression. The recombinant protein was purified and used in an electrophoretic mobility shift assay (EMSA). Indeed, bacterial expressed wt chromobarrel domain directly interacted with a single stranded RNA probe in this experiment, and this interaction was lost upon mutation of Tyr416 or Trp426 (Figure 21B, C).

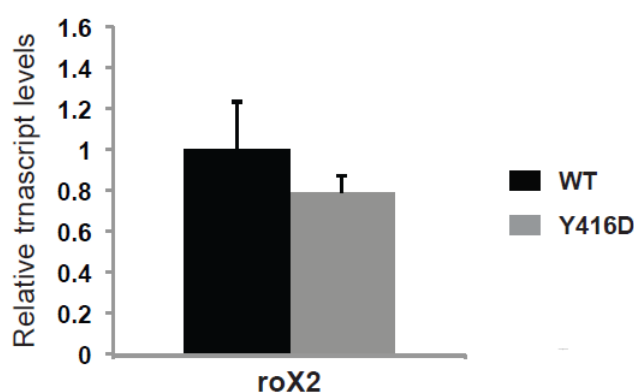


**Figure 21. The MOF chromobarrel domain interacts with RNA, provided by Ibrahim Ilik.**

(A) Ribbon diagram of the *Drosophila* MOF chromo-barrel domain (left panel). Residues Tyr416 and Trp426 are shown as sticks. Structure of the yeast Esa1 chromodomain (residues 17-89, middle panel) in the same orientation as MOF. Structure of the extended chromodomain of yeast Esa1 (residues 1-89, right panel). The N- and C- termini form a short  $\beta$ -sheet and a loop following Tyr56 changes its conformation. (B-C) The MOF chromobarrel domain interacts with RNA. (B) MOF chromobarrel domain (amino acids 346-448, lanes 1-2) and two point mutants (Y416D, lanes 3-4 and

W426G, lanes 5-6) were expressed in BL21 (DE3) cells and purified over Ni-NTA. 1-2µg protein was loaded onto a 4-12% PAA gel and stained with GelCode Blue. (C) EMSA carried out with 25-50-100ng of protein and  $\alpha^{32}\text{P}$ -CTP labeled RNA (84nt, single-stranded) shows that wild-type chromodomain (lanes 2-4) interacts with RNA and forms protein-RNA complexes. In point mutants Y416D and W426G this interaction seems to be completely abolished (lanes 5-10). Lane 1 contains radiolabeled RNA without any protein.

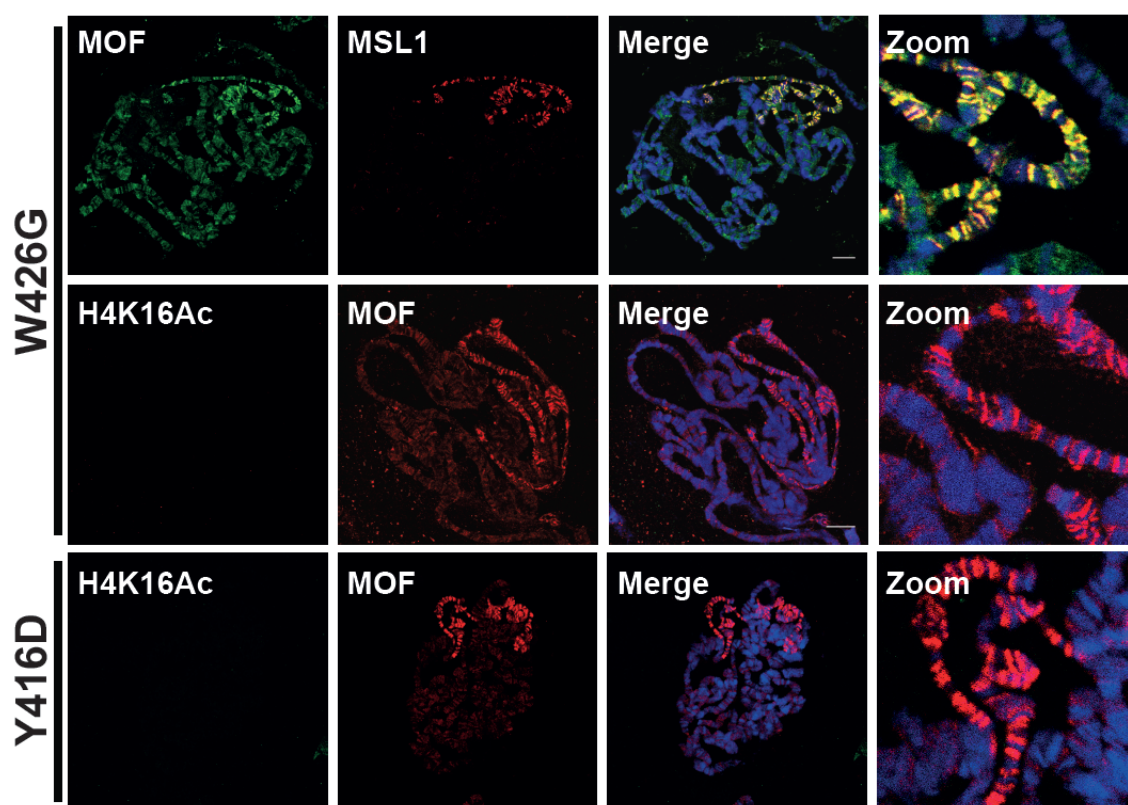
Surprisingly, a later study claimed that the MOF chromobarrel domain had a minor role to play in dosage compensation [231]. However, this study only tested the capacity of MOF to target to the X chromosomal territory by simply overexpressing MOF variants in a male cell line, leaving the possibility that potential defects in MOF function have been masked by the presence of the endogenous protein. Since the chromobarrel domain is required for MOF interaction with roX RNA, we expected to find similar effects on DCC targeting in Y416D and W426G MOF expressing flies as the ones observed in the presence of  $\Delta\text{N}$  MOF, where roX RNA levels are strongly diminished. Strikingly, overexpression of MOF variants carrying point mutations in the chromobarrel domain (Y416D and W426G) failed to rescue male lethality of the *mof2* background (Figure 13 and 14A). This result clearly shows that the essential nature of the MOF chromobarrel domain had been overlooked in previous experimental setups. Intriguingly however, the observed phenotypes differed substantially in further analysis. Upon disruption of the chromobarrel domain, we only found a moderate reduction in roX2 levels *in vivo* (Figure 22).



**Figure 22. Moderate reduction of roX levels upon disruption of the chromobarrel domain.**

roX2 levels in *mof2* male larva salivary glands expressing MOF transgenes as indicated. Transcript levels were normalized to roX2 DNA recovered from the same sample.



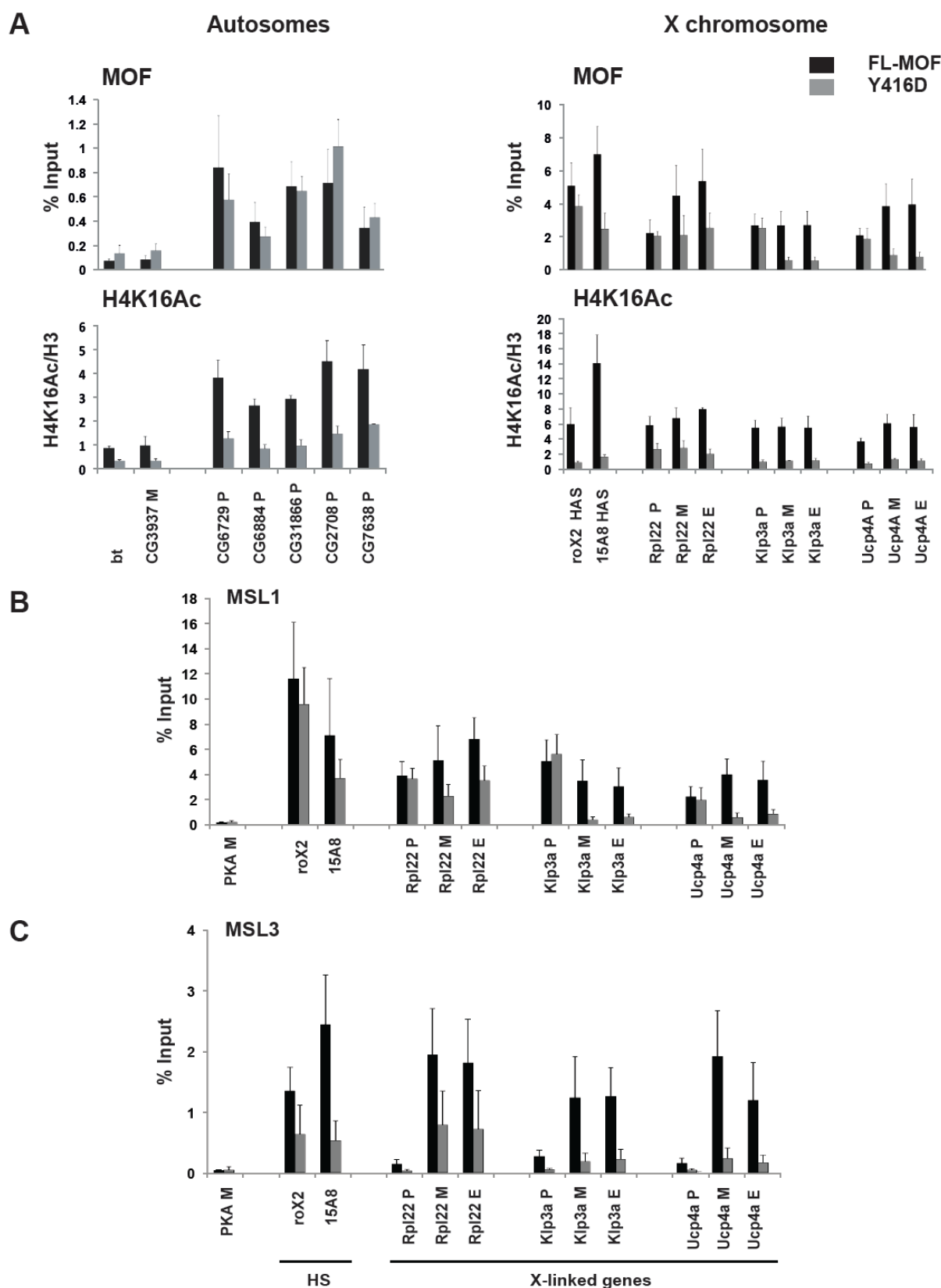


**Figure 23. The chromobarrel domain is required for H4K16ac.**

Immunostaining of polytene chromosomes from *mof2* male third instar larva salivary glands expressing W416G or Y416D MOF as indicated, using antibodies against HA (MOF), MSL1 and H4K16ac. DNA staining is shown in blue (Hoechst 322).

Furthermore, Y416D and W426G MOF were still detected across all autosomes in immunostainings, and appeared enriched on the X chromosome (Figure 23). This result suggested that both chromobarrel domain mutants integrate into the NSL complex as well as the DCC. Indeed, when we immunoprecipitated stably expressed FLAG-tagged Y416D MOF from SL-2 cells, interactions to other MSLs appeared only slightly reduced while interactions to NSL complex members remained unaffected (Figure 17). To our great surprise however, upon immunostaining of polytene chromosomes with antibodies against H4K16ac, we saw a severe reduction of the histone mark across all chromosomes for both chromobarrel domain mutants (Figure 23). This suggested a more general role of the chromobarrel domain than previously anticipated, which seems to involve the activation of MOFs enzymatic capacity after its initial recruitment to chromatin, irrespective of the chromosomal context. Considering the dramatic reduction of H4K16ac upon disruption of the MOF chromobarrel domain, we wanted to assay for defects in chromatin targeting at higher resolution.





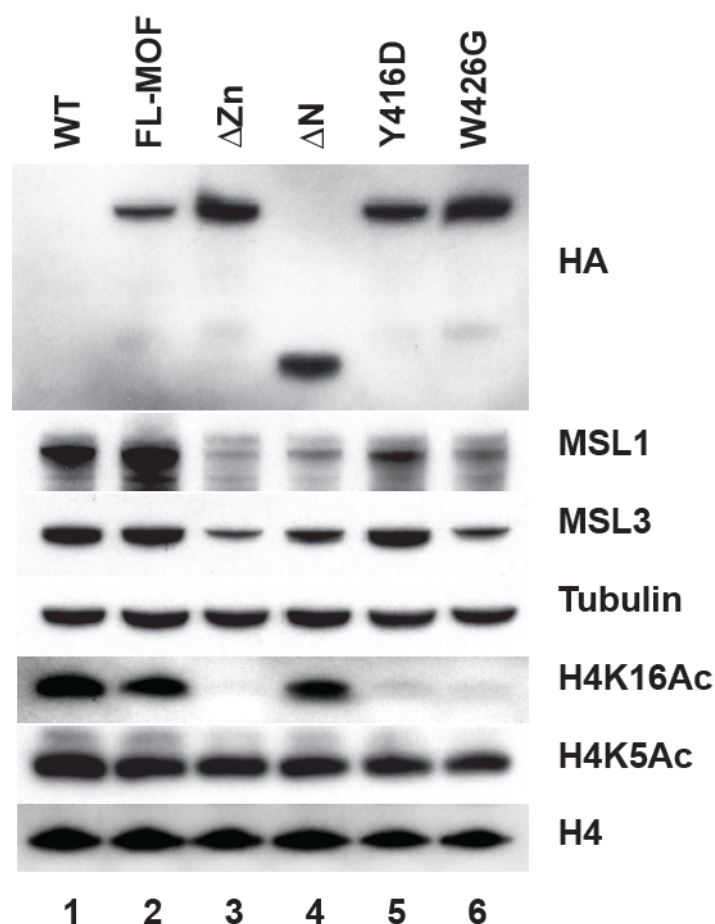
**Figure 24. The chromobarrel domain is required for MSL spreading but not promoter binding.**

(A) MOF targeting is lost from the male X chromosome but retained at autosomal promoters in the presence of Y416D MOF, while H4K16ac is lost globally. ChIP using MOF (upper panel), and H4K16ac (lower panel) antibodies in *mof2* male 3<sup>rd</sup> instar larva expressing FL-MOF (black) or Y416D MOF

transgenes (grey). Binding to the autosomal genes *bt*, *cg3937*, *cg6729*, *cg6884*, *cg31866*, *cg2709*, *cg7638*, the X-linked high affinity sites at the *roX2* gene and location 15A8, as well as the X chromosomal genes *Rpl22*, *Klp3a*, and *Ucp4a* is shown. Primers were positioned at the promoter (P), middle (M), and end (E) of genes. The exact position of the primers is described in the Supplemental Data. ChIP is shown as percentage recovery of input DNA (% Input). Error bars represent standard deviation (StDev) of three independent experiments. (B) MSL1 recruitment is lost from the transcribed region of X-linked genes in the presence of Y416D MOF. ChIP using MSL1 antibodies in *mof2* male 3<sup>rd</sup> instar larva expressing FL-MOF (black) or Y416D MOF transgenes (grey). Binding to the X-linked high affinity sites at the *roX2* gene and location 15A8, the X chromosomal genes *Rpl22*, *Klp3a*, and *Ucp4a*, as well as the autosomal *PKA* gene is shown. Primers were positioned at the promoter (P), middle (M), and end (E) of genes. (C) ChIP as in (B) using antibodies against MSL3.

To this end, we performed ChIP from Y416D MOF expressing male 3<sup>rd</sup> instar larva to monitor differences in MOF binding, H4K16ac and DCC recruitment. Disruption of the chromobarrel domain had diverse effects on MOF targeting. The Y416D mutant MOF protein showed no difference in binding to gene promoters and the *roX2* HAS (Figure 24A). However, binding to the 15A8 HAS and the transcribed region of *Rpl22* was twofold reduced, while targeting to the transcribed region of the low affinity genes *Klp3a* and *UCP4a* was completely lost. This data indicated a specific defect in chromatin targeting of Y416D MOF as part of the DCC, while DCC independent binding to gene promoters remained unaffected. Indeed, when we assayed X-linked target sites for MSL binding, the pattern observed in the presence of Y416D MOF was very similar to the one observed in the *mof2* background, showing a loss of MSL1 and MSL3 from the 3' end of low affinity target genes (Figure 24B). Together, these data confirmed the observation from immunostainings that disruption of the chromobarrel domain impairs DCC targeting to low affinity target sites. We next analyzed H4K16ac across X-linked and autosomal target sites in the presence of Y416D MOF. Strikingly, H4K16ac was strongly reduced across all sites, including gene promoters and HAS, where the Y416D MOF protein was still readily detected (Figure 24A). This result suggests that the chromobarrel domain serves to trigger MOFs catalytic activity after initial recruitment of MOF to its chromatin targets. The observed loss of H4K16ac was striking and surprising. We therefore performed western blots from 3<sup>rd</sup> instar larva extracts to confirm the global reduction in H4K16ac by an independent method. When probing with HA antibodies, we found that all five MOF transgenes were stably expressed to similar levels (Figure 25). Antibodies against tubulin and unmodified histone H4 were used as loading controls. MSL1 levels appeared clearly diminished in the presence of  $\Delta$ Zn,  $\Delta$ N, Y416D and W426G MOF, reflecting compromised MSL complex assembly on X linked genes. MSL3 seemed less

destabilized in the presence of Y416D MOF, but mirrored MSL1 levels in all other mutant backgrounds. FL-MOF efficiently rescued the *mof2*-mediated loss of H4K16ac to wt levels. No H4K16ac was detected upon deletion of the zinc finger in  $\Delta$ Zn MOF, reflecting compromised chromatin targeting of this mutant. Consistent with our previous analysis, substantial amounts of H4K16ac could be detected after deletion of the N-terminus in  $\Delta$ N MOF. Strikingly however, global H4K16ac levels were dramatically reduced in both chromobarrel domain mutants, confirming the general role of this domain for genome-wide H4K16ac. To control for the specificity of the assay we also probed against H4K5ac as well as H4K8ac and H4K12ac. These histone marks are not mediated by MOF in *Drosophila*, and remained unchanged in each of the MOF mutants analyzed (Figure 25 and data not shown).

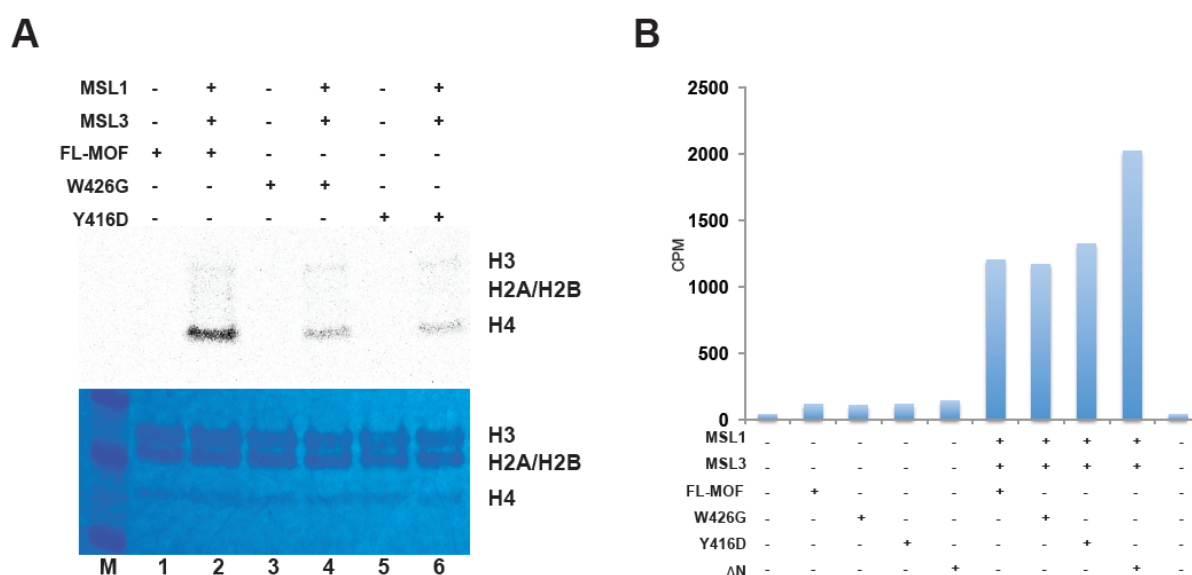


**Figure 25. The chromobarrel domain is required for bulk H4K16ac.**

Western blot analysis of extracts from *mof2* male 3<sup>rd</sup> instar larva expressing the indicated MOF transgenes. Membranes were probed with antibodies as indicated.

### 3.8 The chromobarrel domain triggers acetylation after nucleosome binding

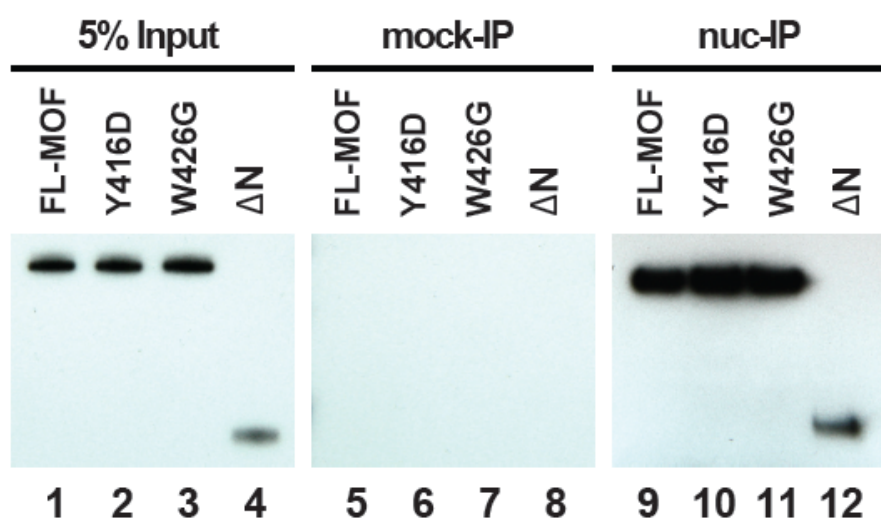
In order to recapitulate the requirement of the chromobarrel domain for MOF HAT activity *in vitro*, we again generated baculoviruses to express and purify Y416D and W426G MOF proteins from SF21 cells and performed HAT assays on nucleosomal substrates. Consistent with our *in vivo* observations, H4 directed acetylation activity was approximately three- to five-fold reduced in trimeric complexes containing Y416D and W426G MOF, as compared to FL MOF (Figure 26A and data not shown). This reduction was specific to nucleosomal substrates, since the same trimeric complexes showed comparable activities towards free histones, which also confirmed functionality of the MOF HAT domain itself (Figure 26B). The activity of  $\Delta N$  MOF containing trimeric complexes towards free histones was slightly enhanced, although the increase was not as pronounced as on a nucleosomal substrate.



**Figure 26. The MOF chromobarrel domain triggers H4K16ac after nucleosome binding, provided by Herbert Holz.**

(A) Acetylation assay on native nucleosomes. Each reaction contains 5.3 nM of FL, Y416D or W426G MOF proteins, with or without MSL1 and MSL3, respectively, and 1.5  $\mu$ g of native nucleosomes purified from MCF-7 cells. (B) Acetylation assay on free histone octamers. 40 ng (21.3 nM) of the indicated recombinant MOF-protein or trimeric complexes were incubated with  $C^{14}$  labeled acetyl coenzyme A and 1.5  $\mu$ g of recombinant (*Xenopus*) histone octamer. The reaction was then applied on a P81 filter paper, air dried, washed and counted in scintillation liquid.

We then wanted to ask at which step of initial substrate binding and subsequent acetylation the chromobarrel domain functions. To address this question we performed affinity purifications of trimeric complexes using biotinylated DNA or mononucleosomes coupled to streptavidin beads. Nucleosome binding of trimeric complexes containing the  $\Delta N$  MOF protein was slightly weakened as compared to FL MOF (Figure 27). This result further highlights the enhanced enzymatic activity of  $\Delta N$  MOF on the same substrate. Strikingly, trimeric complexes containing Y416D and W426G MOF proteins bound as efficiently to the nucleosomal substrate as FL MOF trimeric complexes. Initial binding to the nucleosome is thus independent of a functional chromobarrel domain, which is then required to activate the subsequent acetylation of the H4 tail.



**Figure 27. The chromobarrel domain is not required for nucleosome binding, provided by Herbert Holz.**

Nucleosome pull down experiments are shown using trimeric complexes containing FL, Y416D, W426G or  $\Delta N$  MOF together with MSL1 and MSL3 proteins. Purified endogenous histone octamers from *Drosophila* embryos were wrapped with 147 bp biotinylated DNA to assemble mononucleosomes and coupled to streptavidin beads. After incubation with trimeric complexes and stringent washing, MOF binding was detected by western blot analysis using HA antibodies. 5% of the Input material, IP using empty streptavidin beads as a control (mock-IP), and nucleosome IP (nuc-IP) are shown.



# **Discussion**

#### 4.1 MOF is the major H4K16 specific HAT in *Drosophila*

Since the recent discovery that MOF resides at autosomal gene promoters as part of the NSL complex, the full extent of MOF function at these sites has remained elusive [215, 227, 230]. Disruption of other NSL complex members leads to lethality in males and females [224]. This suggested that MOF is not strictly necessary for NSL complex function, since adult female flies can be recovered in the absence of MOF [186]. However, although MOF was identified in a screen for male specific lethality, this study did not address defects in female viability. We now show that MOF is indeed an essential gene in both sexes, since the number of females reaching adulthood, and especially female lifetime, are drastically reduced in the absence of MOF. Furthermore, the contribution of MOF to H4K16ac at gene promoters had previously remained unclear [230]. Our data demonstrates that, in addition to its role in X chromosome dosage compensation, MOF is targeted to the vast majority of all active gene promoters in male and female flies. Importantly, we also show that in the absence of MOF, H4K16ac is lost from all MOF target sites in the male and female genome. In fact, bulk H4K16ac is lost genome wide in 3<sup>rd</sup> instar larva upon disruption of MOF in both sexes, suggesting that MOF is the major H4K16ac specific HAT in *Drosophila*. However, since a recent study showed that ATAC2 is contributing to bulk H4K16ac during embryonic development [240], the possibility remains that additional enzymes might have the capacity to mediate H4K16ac in certain developmental stages or tissues. Interestingly, bulk H4K16ac is also strongly diminished in human cells upon RNAi of the close homologue hMOF [241, 242], and it has subsequently been shown that hMOF is targeted to thousands of gene promoters across the human genome [243]. It is noteworthy that dosage compensation in mammals does not involve MSL proteins, and accordingly hMOF does not appear enriched on the mammalian X chromosome [243]. The most closely related MYST domain in yeast resides in *S.cerevisiae* SAS2p [237], which hyperacetylates subtelomeric regions to prevent Sir3p binding and spreading of telomeric heterochromatin [244, 245]. However, a better resemblance of the MOF domain architecture is found in the yeast homologue Esa1p, which like MOF contains a chromobarrel domain preceding the MYST domain. Interestingly, Esa1p is generally recruited to the promoters of active protein coding genes where it mediates acetylation of histone H4 as part of the NuA4 complex, and this complex also contains a yeast homologue of *Drosophila* MSL3, Eaf3 [246]. It is therefore highly likely that the pattern of MOF binding that we observe at promoters across male and female autosomes and on the female X chromosome in *Drosophila*, is indeed reflecting the most ancient mode of MOF function.

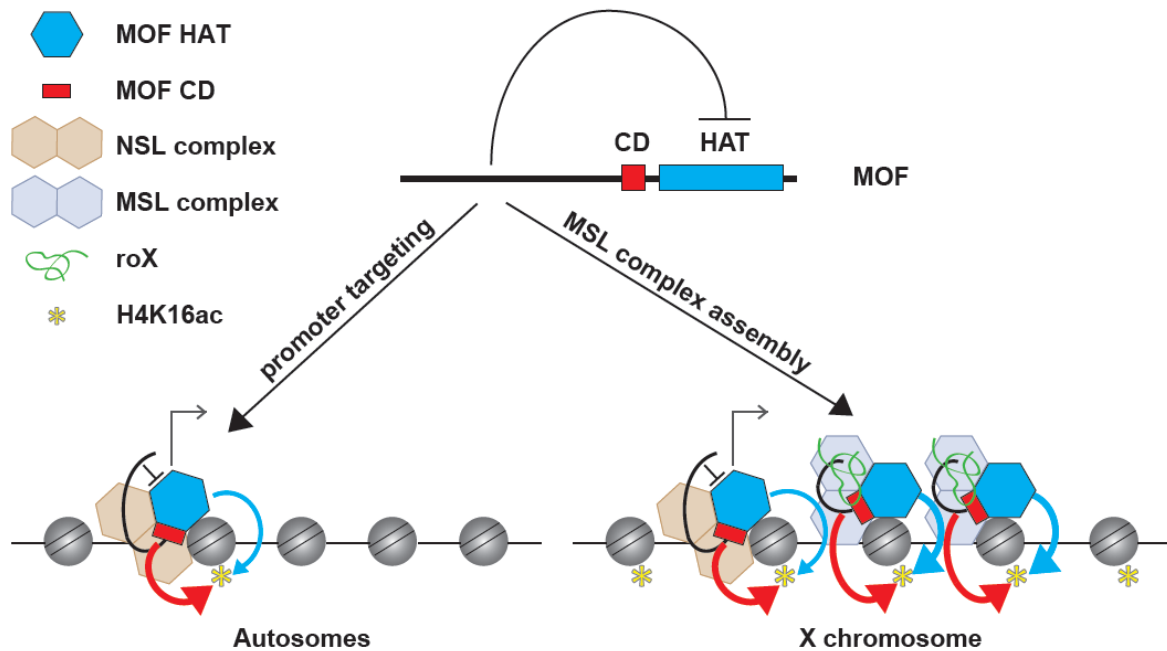


## 4.2 MOF controls dosage compensation via its N-terminal domain

A prominent feature of MOF in all *Drosophila* species is the presence of a large N-terminal region, which is absent in all other MOF homologues from yeast to human. Addition of this domain correlates with the evolution of the *Drosophila* dosage compensation system, involving the targeting of MSLs and H4K16ac to the male X chromosome [247]. It seems thus plausible that the N-terminal domain has added functions that utilize the evolutionarily ancient transcriptional regulator MOF for the novel task of dosage compensation. Strikingly, upon deletion of the MOF N-terminus, MSL targeting to the X chromosome was completely abolished. This also included high affinity sites (HAS), which are otherwise resistant to MOF depletion [185]. Deletion of the MOF N-terminus therefore has a dominant negative effect on DCC formation. Aberrant MSL targeting was reminiscent of the phenotypes observed in roX1 and roX2 double mutants [191], and indeed roX levels were reduced by 98% in the absence of the MOF N-terminus. This result suggests that, although interactions to MSL1 and MSL3 are mediated by the zinc finger region of MOF, the N-terminus is required to assemble the core MSL subunits together with roX RNAs into a functional DCC. At the same time, the N-terminus is also required for proper integration of MOF into the NSL complex at gene promoters. Importantly, despite these defects, N-terminally truncated MOF was still active as a HAT and able to target to chromatin (Figures 14 and 16). In marked contrast to this, disruption of the evolutionarily conserved zinc finger, which is a feature common to all HATs of the MYST type, completely disrupted MOF function (Figures 19 and 20).

MSL protein domains required for DCC assembly and targeting, as well as their cognate DNA binding sequences have been shown to undergo rapid adaptive coevolution in *Drosophila melanogaster* [248, 249]. Also roX RNA sequences are highly divergent throughout *Drosophila* species [238]. Likewise, the N-terminal domain of MOF shows huge variation in size and amino acid sequence between *Drosophila* species suggesting ongoing selective pressure by other MSLs or roX RNA. Another striking feature of the N-terminal domain is its intrinsic disorder, according to secondary structure prediction. It is a recently emerging concept that protein function is not necessarily linked to fixed secondary and tertiary structures [250]. Functional disordered regions have been particularly suggested for domains in hub proteins that control a variety of biological processes and mediate interactions to multiple interaction partners [251]. The most prominent example of this type of regulation is p53, which interacts with hundreds of binding partners via intrinsically unstructured domains. The N-terminal domain of MOF has relatively few hydrophobic amino acids, suggesting solubility in solution. We therefore propose that the unstructured N-terminal

domain integrates the multiple functions of MOF in dosage compensation and for genomewide H4K16ac (Figure 28).



**Figure 28. Several layers of regulation control genomewide H4K16ac in *Drosophila*.**

Our data suggests that in *Drosophila* MOF is responsible for the majority of H4K16ac at gene promoters in males and females, and for hyperacetylation along the male X chromosome. MOF activity is tightly regulated to achieve these diverse tasks. The Nterminal region is required for targeting of MOF to gene promoters and for assembly the MSL complex. At the same time the N-terminus constrains the enzymatic activity of the MOF HAT domain, which is only unleashed in the context of the MSL complex to achieve hyperacetylation of the male X (blue arrows). The next layer of regulation is imposed by the chromobarrel domain (CD). Following binding of MOF to its chromatin target sites, nucleic acid interactions of the chromobarrel domain are required to trigger acetylation of the H4 tail (red arrows), which subsequently allows MSL complex spreading along the transcribed regions of genes. The chromobarrel domain thus acts as an “on-off switch” for H4K16ac.

A further intriguing finding was that the N-terminus constrains MOF HAT activity *in vivo* and *in vitro*. Again the necessity for this additional level of regulation arises from the dual function that MOF has adopted in *Drosophila*. We show that H4K16ac displays much higher baseline levels on the male X chromosome as in any other chromosomal context. Furthermore, MOF mediated H4K16ac extends beyond regions of MSL binding on the male X chromosome. We do not observe the same phenomenon on male autosomes or in females, where much lower levels of H4K16ac are restricted to sites of MOF binding at gene promoters. The activity level of MOF thus differs to a large degree depending on the

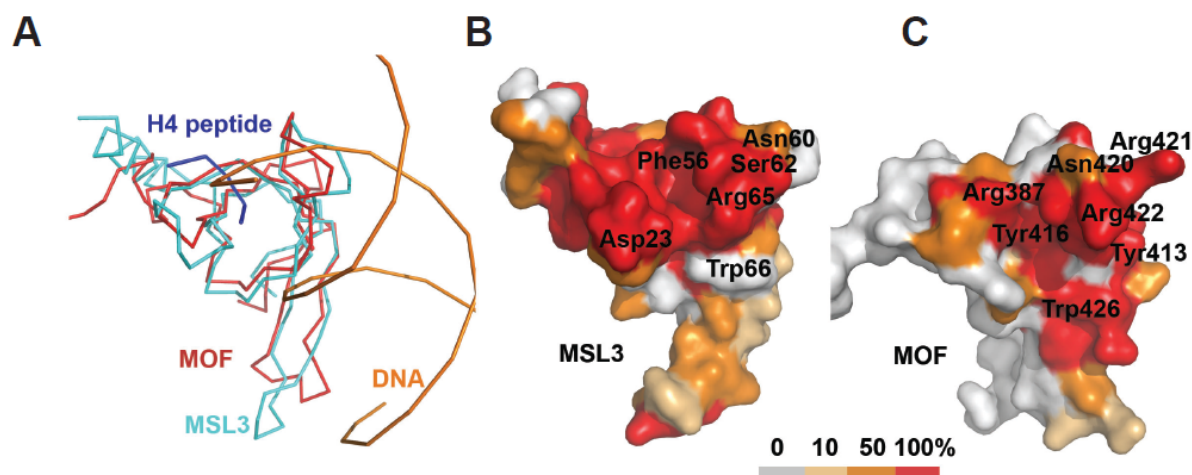
chromosomal context. It is known that MOF requires interaction with MSL1 and MSL3 to be active as a HAT *in vitro*. However, our data suggests that an additional constrain is imposed on MOF's enzymatic activity by an autoinhibitory function residing in its N-terminus. Interestingly, our data shows that in the  $\Delta N$  MOF background, H4K16ac spreads from sites of MOF binding into neighboring regions even in the absence of other MSLs. We therefore propose that in the wildtype situation, the N-terminus may control or restrict the H4K16ac spreading around sites of MOF binding, and this constrain is only released in the presence of a fully assembled DCC on the male X chromosome for more extensive acetylation (Figure 28). In previous experimental setups, loss of MOF mediated H4K16ac has been accompanied by simultaneous loss of MSLs from the transcribed regions of X-linked genes [185, 186, 215, 230]. Therefore the possibility remained that H4K16ac could merely serve to allow MSL spreading along the X chromosome, and upregulation of transcription would be achieved subsequently by additional activities residing in other MSL proteins. Although MSLs were entirely lost from the body of low affinity target genes in the absence of the MOF N-terminus, the reduction in H4K16ac was less pronounced, most likely reflecting the presence of hyperactive MOF at nearby promoters. Intriguingly, about 11% of males escaped lethality in this background. This suggests that in these individuals, the presence of H4K16ac on X chromosomal genes has been sufficient to ensure dosage compensation despite the lack of MSL binding. This result supports the view that H4K16ac is sufficient to upregulate transcription of X-linked genes, and that the main function of the DCC is to direct H4K16ac to the male X chromosome.

#### 4.3 The MOF chromobarrel domain controls H4K16ac genomewide

An earlier study claimed that the MOF chromobarrel domain had a minor role to play in dosage compensation [231]. However, this study only tested the capacity of MOF to target to the X chromosomal territory by simply overexpressing MOF variants in a male cell line, leaving the possibility that potential defects in MOF function have been masked by the presence of the endogenous protein. Indeed, demonstrating the advantages of the *in vivo* system, we were able to reveal the crucial role that the MOF chromobarrel domain plays for all aspects of MOF function. Upon disruption of its nucleic acid binding properties, MSL spreading to X-linked genes is compromised, leading to a defect in dosage compensation. This is reminiscent of the phenotype observed upon deletion of the chromobarrel domain of MSL3 [212]. However, the MSL3 chromobarrel domain is thought to contribute to MSL

targeting via its binding to the H3 tail trimethylated at K36 [210]. Also an interaction with the H4 tail monomethylated at K20 has been proposed [213, 214]. In contrast, the MOF chromobarrel domain lacks the aromatic cage required for binding to methylated lysine residues. We therefore believe that reduced MSL spreading in MOF chromobarrel domain mutants is not the result of a direct chromatin binding defect of the MOF protein, consistent with the fact that binding of MOF to gene promoters *in vivo* and to nucleosomes *in vitro* is unaffected upon disruption of the chromobarrel domain. It has been shown previously that the chromobarrel domain mediates MOF binding to roX RNAs, and it is possible that this interaction contributes to spreading of the DCC along dosage compensated genes. However, the most dramatic and unexpected consequence upon disruption of the chromobarrel domain was the dramatic loss of genomewide H4K16ac. It is clear from previous work that this reduction of H4K16ac alone is sufficient to disrupt MSL spreading on the X chromosome. Indeed, upon direct mutation of the catalytic site in the MOF enzyme, MSL binding is restricted to high affinity sites in a pattern similar to the one observed in chromobarrel domain mutants [185].

The structure of the *Drosophila* MOF chromobarrel domain (aa 367-454) has been determined by NMR [235] (Figure 21A). While Tyr416 (Y416) is one of the putative aromatic cage residues and is partially buried in the core of the structure, Trp426 (W426) is solvent exposed on a  $\beta$ -sheet formed by strands  $\alpha 2$ ,  $\alpha 3$  and  $\alpha 4$ . As mentioned above, a similar domain is present also in the yeast Esa1p. It has been shown there that in presence of a short N-terminal extension, a minor conformational change occurs in the core of the Esa1p domain, triggering its RNA/DNA binding activity [252]. Interestingly, a similar structure of the chromobarrel domain of MSL3 has recently been determined in complex with dsDNA and the N-terminal tail of histone H4 monomethylated at H4K20 [213]. Both MOF and Esa1 possess equivalent highly conserved surfaces that could be involved in the interaction with nucleic acids (Figure 21A). Accordingly, extensive mutagenesis studies in Esa1p identified residues involved in nucleic acid binding, including residues corresponding to MOF Tyr413, Tyr416, Asn420, Arg422 and Trp426. Mutations of these residues in Esa1p were lethal or produced severe growth effects [252]. In the MSL3 structure, Trp66 corresponding to MOF chromobarrel domain Trp426 is directly involved in the interaction with DNA (Figure 29). Although the aromatic residues that are critical for interaction of the MSL3 chromobarrel domain with the methylated lysine side chain of the H4 peptide are not well conserved, it remains unclear whether MOF and Esa1 could bind unmodified histone tail residues.



**Figure 29. Conservation between the MOF and MSL3 chromobarrel domains, provided by Jan Kadlec.**

(A) Ribbon representation of the complex of MSL3, DNA and the H4 peptide (PDB code 3OA6). The structure of the MOF chromobarrel domain (PDB code 2BUD) is superimposed onto the one of MSL3. (B) Surface representation of the MSL3 chromobarrel domain in the same orientation as in (A) highlighting areas of conserved residues involved in DNA and H4 binding. The conservation of the surface is represented from gray to red (red is 100 % conserved) according to the color scale bar. The aromatic cage around Phe56 includes also Tyr31, Trp59 and Trp63. (C) Surface conservation of the MOF chromobarrel domain based on ortholog sequences ranging from human to nematodes. The conserved surface corresponding to the MSL3 DNA binding region includes Asn420, Arg422, Tyr314, Trp426 and Arg387. Tyr416 forms a floor of a cavity surrounded by Leu419, Leu423 and His393.

However, one could potentially envisage that a mode of action similar to MSL3 may also exist for the Esa1p and MOF chromobarrel domains, and that interaction with nucleic acid and the unmodified H4 tail direct the tail for acetylation by the MYST domain. It has been shown previously that the MOF chromobarrel domain binds to nucleic acids with a preference for RNA over DNA, and interacts with roX RNAs *in vivo* [236]. We now show that the nucleic acid binding properties of the MOF chromobarrel domain are necessary to trigger H4K16ac on a nucleosomal substrate *in vitro*, where the local DNA concentration is high due to tethering of the complex to the nucleosome via other domains. Interestingly, deletion of a conserved stem loop structure in roX RNAs leads to a specific loss of hyperacetylation from the male X chromosome *in vivo*, while MOF is still targeted to the X chromosome in this background [238, 239]. It is therefore tempting to speculate that upon ordered assembly of roX RNAs and MSL proteins *in vivo*, roX RNAs could engage in similar interactions with the chromobarrel domain and H4 tail to promote high acetylation levels throughout the body of X-linked genes. Since the chromobarrel domain is also conserved in

human MOF it is likely that MOF's acetylation activity is controlled in a similar manner in humans.

#### 4.4 Summary

In this study we have revealed the function of the MOF chromobarrel domain, which is to elicit the activity of its associated HAT enzyme on the chromatin target. The huge degree of sequence conservation of the MOF protein between flies and mammals suggests that a similar mode of operation might be present in mammalian MOF. Our findings thus have important implications for the study of this enzyme in humans, where it is implicated in a wide range of processes like transcription, DNA repair and cancer [121, 243, 252]. In addition, it will be exciting to test if chromodomains in other enzymatic complexes can directly control their associated enzyme activities in a similar way. In contrast to the chromobarrel domain, the N-terminal part of the MOF protein is specific to *Drosophila* species and regulates MOFs function in dosage compensation. Accordingly, our work also highlights how a MYST HAT has been adopted through evolution to carry out distinct tasks on X chromosomal and autosomal genes. We propose that the enzymatic activity of other MYST family HATs may be under similar regulation by associated domains for context specific function.

## **Ongoing Projects**

### 5.1 The mechanism of dosage compensation

As I have pointed out before, the potentially most important question in the field of *Drosophila* dosage compensation is at what level transcriptional upregulation of male X-linked genes is achieved. Identifying the step in the RNA PolII transcription cycle that is targeted by MSL mediated transcription activation is the first step towards unraveling the mechanism of dosage compensation. During my PhD, I developed a ChIP protocol of extremely high sensitivity. I used this method to generate duplicate ChIP-Seq profiles of the Rpb3 subunit of RNA PolII from approximately 3000 pairs of 3<sup>rd</sup> instar larva salivary glands from wt male, wt female and male MSL2 RNAi animals. Table 2 shows the number of sequencing reads obtained for each data set.

Sample name	# reads	# reads mapped	% reads mapped	coverage
<b>Male samples:</b>				
polII maleA1 wt	25,556,049	19,270,618	75.41	8.68
input maleA1 wt	25,936,578	20,906,815	80.61	9.42
polII maleA2 wt	24,802,445	20,376,581	82.16	9.18
input maleA2 wt	24,657,849	20,392,877	82.70	9.19
<b>Female samples:</b>				
polII femaleA1 wt	24,502,071	18,696,145	76.30	8.42
input femaleA1 wt	29,133,963	23,527,559	80.76	10.60
polII femaleA2 wt	26,164,099	17,055,582	65.19	7.68
input femaleA2 wt	25,881,776	21,181,705	81.84	9.54

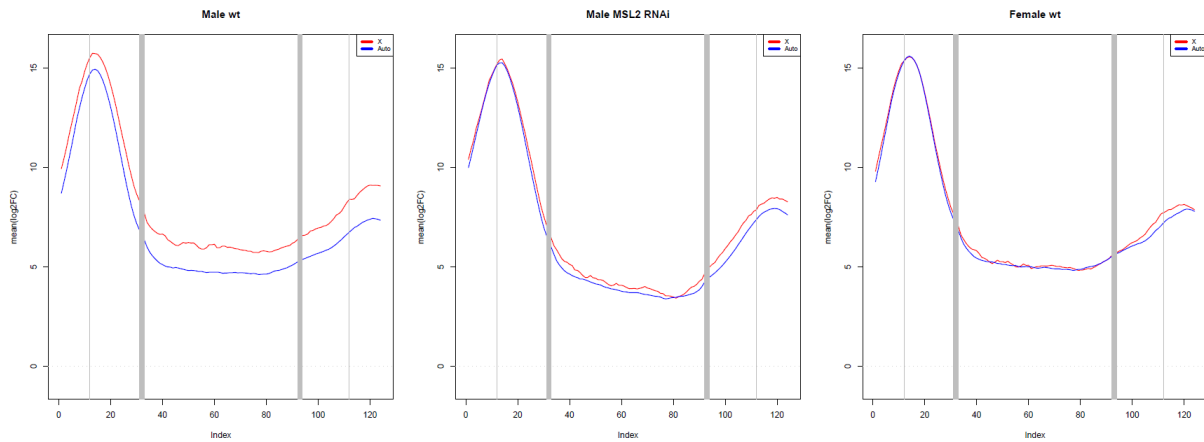
**Table 2: ChIP-Seq samples with numbers and percentages of mapped reads.**

This table shows the total numbers of reads sequenced from each sample, the numbers and percentages of the reads mapped on the *D. melanogaster* genome by Bowtie [253], and the genome coverage obtained from these mapped reads.

The bioinformatics analysis of this data, performed by my collaborators at the European Bioinformatics Institute, was still ongoing at the time of submission of this thesis, and experiments to support our early conclusions are currently underway. For this reason I will just give a short outlook into this ongoing project. Nevertheless, the data obtained so far already enables us to make some important conclusions. Figure 30 shows the resulting average plots of PolII on X-linked and autosomal genes. It is immediately evident that the levels of PolII are approximately two fold enriched in the transcribed region of male X-linked



genes, demonstrating that our method is able to accurately capture enhanced transcription by PolII.



**Figure 30: Differences in Pol II binding between dosage-compensated and non-compensated genes.** (A) Average Pol II-binding profiles of X-linked (red) and autosomal (blue) genes are shown for males, females and male MSL2 RNAi. Grey lines indicated the transcription start site and 3' end of genes. An unscaled window was used for the promoter region and end of genes; (TSS-300bp, TSS+500bp) and (3'end-500bp, 3'end+300bp). The transcribed region of genes (TSS+500bp, 3'end-500bp) was scaled according to gene length.

This result is in agreement with a recent study that showed a similar enrichment in the transcribed region of X-linked genes using the GRO-Seq (global run on) method, albeit with lower sensitivity [217]. However, one major conclusion of that study, which was limited to a male cell line system and did not include a female data set, was that dosage compensation does not affect the levels of RNA PolII at gene promoters. Strikingly, in our analysis we find that males show an increase in the binding of PolII to X-linked relative to autosomal promoters, when compared to the female situation. Importantly, increased promoter binding was lost upon MSL2 RNAi in males. The magnitude of the change is roughly two-fold ( $\sim 0.85$  and  $\sim 0.95$  in  $\log_2$  scale for the comparison male vs. female and male vs. MSL2 RNAi respectively). Our data therefore strongly suggests that dosage compensation in *Drosophila* is at least partially mediated by enhanced transcription initiation on the male X chromosome. The hall mark chromatin modification of the male X chromosome, H4K16ac, is known for its ability to create an open and permissive chromatin structure, as I have elaborated on in the introduction of this thesis. It seems therefore convincing that increased access to the transcription machinery and associated factors leads to enhanced transcription initiation and thus a two fold increase in gene expression. Accordingly, MSL binding along the body of X-linked genes might serve to spread acetylation by MOF, in order to create this permissive

environment. H4K16ac in the body of X-linked genes might also serve to pave the way for the increased amount of PolII loaded onto these genes in males. Global ChIP-Seq analysis does not allow the direct comparison of independent data sets. Therefore we cannot conclude at this point that the shift in rank order of male X-linked promoters is due to an increase of PolII binding at X-linked promoters or due to a decrease at autosomal promoters. To clarify this point by an absolute measurement of PolII binding in males females and upon MSL2 RNAi, and to verify enhanced initiation on the male X chromosome, I will next perform ChIP against Rpb3, PolII phosphorylated at Serine5 and TATA binding protein (TBP) and determine the binding at a set of X-linked and autosomal target genes as absolute Input recoveries. A further conclusion of the above mentioned study by the Kuroda lab was that autosomal genes generally suffer from decreased PolII processivity, manifested by a decreasing slope of PolII across the transcribed regions of autosomal genes [217]. We were not able to reproduce this result, since the slope of PolII across transcribed regions appeared similar on X-linked and autosomal genes in males and females, although a decrease in processivity was observed on all chromosomes upon MSL2 RNAi.

Here, some fundamental limitations of the study presented by the Kuroda lab should be pointed out. Firstly, the conclusion that dosage compensation is not acting at the step of transcription initiation was based on the observation that PolII density is similar at X-linked and autosomal promoters in male S2 cells. However, the authors did not obtain PolII profiles from female flies or cell lines to investigate sex specific differences in transcription initiation. Secondly, a major concern regarding the Kuroda study is the use of the S2 cell line for GRO-Seq. It has been shown previously that S2 cells show a high degree of aneuploidy, and that an MSL independent dosage compensation system is operating in these cells to equalize expression between different genomic loci despite high copy number variation (Zhang et al., 2010), an observation that has also been confirmed in vivo [225]. Two important facts stick out from the work by Zhang et al., suggesting that S2 cells are generally unsuitable for the analysis of X chromosome compensation. First, aneuploidy is unevenly distributed between chromosomes, with 40% of autosomal regions showing variations in copy number but only 18% of X-linked loci. Second, the MSL independent compensation mechanism shows less activity towards the male X chromosome. Any analysis aimed at determining engaged PolII in these cells is therefore prone to detect the signature of MSL independent compensation of genomic aneuploidies, with a strong autosomal bias. In contrast to this, our study is based on the analysis of euploid genomes, comparing the same set of genes in the male and female

situation. We are therefore confident that our study will provide important insights into the mechanism of dosage compensation in *Drosophila*.



## **Published Work**

## 6. Contributions to other studies

In addition to the unpublished work presented in the first part of this thesis, I contributed significantly to two major publications during the course of my PhD:

**“Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in *Drosophila melanogaster*”**. Mol Cell Biol 28, 3401-3409.

In the context of this study, I showed that HypB/Set2, the enzyme responsible for trimethylation of H3K36 in *Drosophila*, is localized to active genomic regions of polytene chromosomes and partially colocalizes with MSL1 on the male X chromosome. I furthermore showed by RNAi of HypB/Set2 that removal of the H3K36me3 mark leads to defects in dosage compensation. I also contributed to the writing of the manuscript.

**“The nonspecific lethal complex is a transcriptional regulator in *Drosophila*.”**  
Mol Cell 38, 827-841.

In the context of this study, I showed by immunostainings of polytene chromosomes that NSL recruitment to autosomal targets is independent of MOF. I also show that the transcriptional changes of autosomal genes upon removal of MOF are of less magnitude than the ones observed upon depletion of NSL proteins. Furthermore, I made a major contribution to the development of the ChIP protocol that is used throughout this study. In addition, I partially contributed to the writing of the manuscript.

Both original manuscripts can be found in the Appendix.

## **Materials and Methods**

## 7.1 Chromatin immunoprecipitation (ChIP)

### 7.1.1 ChIP from salivary glands

100 Inverted 3rd instar larva were fixed in fixing solution (50 mM HEPES pH 7.6, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 3.6% formaldehyde) for 20 min at room temperature on a rotating wheel. Formaldehyde was quenched with stop solution (PBS, 0.01% Triton X-100 0.125 M glycine) for 5min. Fixed larva were washed for 5 x 2 min with buffer A (10 mM HEPES pH 7.6, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100) and 5 x 5 min with buffer B (200 mM NaCl, 10 mM HEPES pH 7.6, 1 mM EDTA, 0.5 mM EGTA, 0.01% Triton-X 100). Salivary glands were dissected into 500  $\mu$ L of RIPA buffer (25 mM HEPES pH 7.6, 150mM NaCl, 1 mM EDTA, 1% Triton-X 100, 0.1% SDS, 0.1% DOC, protease inhibitors) and sonicated 16 x 30 sec using a Branson 250 sonicator at 40 pulse, intensity 5, generating 200 bp fragments on average. The chromatin was cleared by 10 min of high speed centrifugation at 4°C in a table top centrifuge. The chromatin was then aliquoted in 8 samples and snap frozen. For immunoprecipitation, one aliquot was filled up with RIPA buffer to 600 $\mu$ l and incubated with 2 $\mu$ l of rabbit anti-MOF, 2  $\mu$ l rabbit anti-H4 (abcam ab10158), or 5 $\mu$ l rabbit anti-H4K16ac (Santa Cruz sc-8662-R) antibodies O/N at 4°C. 50  $\mu$ l of this material was then taken as Input control and 500  $\mu$ l used for immunoprecipitation. For ChIP-Seq, the complete material from 4 x 100 salivary glands was used for Immunoprecipitation, using 3 $\mu$ l of rabbit anti-MOF, 3  $\mu$ l rabbit anti-H4, 5 $\mu$ l rabbit anti-H4K16ac, or 6 $\mu$ l of rabbit anti-Rpb3 (gift from John Lis and Karen Adelman). Immunocomplexes then isolated by adding protein A -Sephadex (Roche) for 3 hours, followed by six washing steps: 4x RIPA buffer, 1x DOC buffer (10 mM Tris at pH 8, 0.25 M LiCl, 0.5% NP-40, 0.5% DOC, 1 mM EDTA), and 1x TE at pH 8. The beads were resuspended in 90  $\mu$ l TE and, together with the Input control, the crosslink was reversed at 65°C over night. After 30 min incubation at 37°C with RNaseA (0.2 mg/ml), followed by 2 hr Proteinase K digestion (0.05 mg/ml) at 50°C, DNA was purified using Minelute columns (Qiagen). ChIP DNA samples were resuspended in 500  $\mu$ l nuclease free water. We used 10  $\mu$ l ChIP material for each qPCR reaction. For ChIP-Seq, the material from four independent IPs was pooled for each experiment (two replicates were generated for all Rpb3 data sets).

### 7.1.2 ChIP from whole larva

For whole larva ChIP, for each replicate, 100 larva were crushed to a powder in liquid nitrogen, dounced 30x in 20 ml NE buffer (15mM Hepes pH 7.6, 10mM KCl, 5mM MgCL<sub>2</sub>,



0.1mM EDTA, 0.5mM EGTA, 350mM sucrose, 0.1% Tween, 1mM DTT, protease inhibitors) and fixed by adding 1ml 36% formaldehyde to 1.8% final concentration for 20 min at RT. After quenching for 5 min at RT with 125 mM glycine, nuclei were collected by 10 min centrifugation at 2500g, washed 3x 5min in RIPA (25 mM HEPES pH 7.6, 150mM NaCl, 1 mM EDTA, 1% Triton-X 100, 0.1% SDS, 0.1% DOC, protease inhibitors) sonicated 5x 30 sec in a Branson 250 sonicator at 30 pulse, intensity 3, followed by sonication in a Covaris SL-2 sonicator for 6 min using the preset 200bp program. After clearing by high-speed centrifugation, the chromatin was split in 8 aliquots and snap frozen for subsequent immunoprecipitation. For immunoprecipitation, the volume was made up to 600  $\mu$ l with RIPA. 2  $\mu$ l of rabbit MSL1 or MOF, 3 $\mu$ l of rat MSL3, 5  $\mu$ l of anti-H4K16ac (Santa Cruz sc-8662-R) 2  $\mu$ l anti-H4 (abcam ab10158), 2  $\mu$ l anti-H3 (ab1791) antibodies were added and incubated over night at 4°C, respectively. Immunocomplexes were isolated by adding protein A/G-Sepharose (Roche) for 3 h, followed by six washing steps: 4x RIPA buffer, 1x DOC buffer (10 mM Tris at pH 8, 0.25 M LiCl, 0.5% NP-40, 0.5% DOC, 1 mM EDTA), and 1x TE at pH 8. The beads were resuspended in 90  $\mu$ l TE and, together with the Input control, the crosslink was reversed at 65°C O/N. After 30 min incubation at 37°C with RNaseA (0.2 mg/ml), followed by 2 hr Proteinase K digestion (0.05 mg/ml) at 50°C, DNA was purified using Minelute columns (Qiagen). ChIP DNA samples were resuspended in 500  $\mu$ l nuclease free water. We used 10  $\mu$ l ChIP material for each qPCR reaction.

## 7.2 RNA isolation and quantitative PCR analysis

For qRT-PCR, RNA and corresponding genomic DNA from SL-2 cells or from salivary glands were isolated simultaneously using the AllPrep DNA/RNA mini kit (QIAGEN), the RNA column was DNaseI treated, and 300 ng of total RNA was used in an RT reaction. qPCR was performed on an Applied Biosystems (AB) Cyclor7500 with SYBR detection, and the amplification curves were analyzed with the corresponding AB software. Each qRT-PCR was repeated at least three times, values were normalized to corresponding genomic DNA values, and the standard deviation within each experiment was calculated. qPCR analysis of ChIP samples was performed using the SYBR Green PCR master mix (Applied Biosystems), 100 ng of each forward and reverse primer, and 10  $\mu$ l immunoprecipitated DNA, in an ABI7500 real-time PCR thermocycler (Applied Biosystems, Inc.). Recovery was determined as the amount of immunoprecipitated DNA relative to input DNA. Error bars represent standard deviation (StDev) of four independent experiments.

### 7.3 Fly stocks and crosses

All stocks were maintained on standard medium at 25°C. To assay for female viability we conducted the cross  $y/mof^2;sb/P\{w+ UAS-HA-MOF\} \times mof^2/mof^2;P\{w+ UAS-HA-MOF\}/tm3$  and the control cross  $y/w-;sb/+ \times w-/w-;+/tm3$  and compared the proportion of  $mof^2/mof^2;sb/t3$  or  $w-/w-;sb/t3$  in the respective offspring. For the complementation test, female flies of the genotype  $mof^2/fm7;P\{w+ UAS-HA-MOF\}$  were crossed to  $y/fm7;P\{armadillo-GAL4\}$  males to induce transgene expression, and the ratio of male  $Y/mof^2$  to female  $mof^2/fm7$  offspring was scored as relative male survival.

For the in vivo characterization of MOF variants the following fly strains were generated:

1.  $w; P\{w+ UAS-HA-MOF\}/TM3$
2.  $w; P\{w+ UAS-HA-\Delta565-587-MOF\}/TM6$
3.  $w; P\{w+ UAS-HA-\Delta565-587-MOF\}/TM3$
4.  $w; P\{w+ UAS-HA-\Delta1-349-MOF\}/Cyo$
5.  $w; P\{w+ UAS-HA-\Delta1-349-MOF\}/TM3$
6.  $w; P\{w+ UAS-HA-\Delta1-349-MOF\}/TM3$
7.  $w; P\{w+ UAS-HA-Y416D-MOF\}/Cyo$
8.  $w; P\{w+ UAS-HA-Y416D-MOF\}/Cyo$
9.  $w; P\{w+ UAS-HA-Y416D-MOF\}/TM6$
10.  $w; P\{w+ UAS-HA-W426G-MOF\}/TM3$
11.  $w; P\{w+ UAS-HA-W426G-MOF\}/Cyo$
12.  $w; P\{w+ UAS-HA-\Delta176-228-MOF\}/$
13.  $w; P\{w+ UAS-HA-\Delta176-228-MOF\}/$
14.  $w; P\{w+ UAS-HA-\Delta241-357-MOF\}/$
15.  $w; P\{w+ UAS-HA-\Delta241-357-MOF\}/$
16.  $w; P\{w+ UAS-HA-\Delta98-357-MOF\}/Cyo$
17.  $w; P\{w+ UAS-HA-\Delta98-357-MOF\}/Cyo$

### 7.4 Immunostaining of polytene chromosomes and confocal microscopy

Two to three pairs of salivary glands were dissected in PBS for each squash. The glands were transferred into a drop of 35  $\mu$ l Fix solution (1.8% PFA, 45% Acetic acid, in H<sub>2</sub>O) on a glass slide and incubated for 10 min. A cover slip was put on top of the sample and tapped with a pencil to disrupt cells and nuclei, then heavily pressed with the thumb to spread the

chromosomes. The slide was snap frozen in liquid nitrogen, the cover slip removed with a razor blade and the slide transferred into PBS. The slide was then incubated in blocking solution (3% BSA, 0.2% NP40, 0.2% Tween20, 10% dry milk, in PBS) for 1 h. After transfer into a wet box, 40  $\mu$ l blocking solution with primary antibody were added onto the sample and incubated for 1 h at RT or over night at 4°C. MOF and MSL1 antibodies were used at 1:500, anti-K16ac Santa Cruz at 1:125 dilutions. After two times rinsing with PBS, the slide was washed for 15 min at RT with washing buffer (500mM NaCl, 0.2% NP40, 0.2% Tween20, in PBS). The slide was then incubated in the dark with 40  $\mu$ l blocking solution containing a 1:250 dilution of fluorescently labeled secondary antibody. The slide was rinsed two times with PBS and washed for 15 min at RT with washing buffer. After another rinse with PBS, 15 $\mu$ l of Fluoromount were added onto the sample and the sample was covered with a cover slip. Images were captured with an AxioCamHR CCD camera on a Leica SP5 (Leica Microsystems) using an Apochromat NA 1.32 oil immersion objective. Images were arranged with Adobe Illustrator.

### **7.5 Generation of protein extracts and western blotting.**

To determine protein levels in adult fly heads, flies were first anesthetized under CO<sub>2</sub> and then sorted by sex. 50 flies were selected, transferred into an eppendorf tube and snap frozen in liquid nitrogen. Upon vigorous shaking of the tube, heads detach and can be collected into 100 $\mu$ l of 2 x Laemmli buffer. Heads were sonicated for 10 sec in a Branson 250 sonicator at 30 pulse, intensity 3. After 10 min centrifugation at full speed in a table top centrifuge, the supernatant was collected and boiled for 5 min at 95°C. 10 $\mu$ l of the sample were then used for PAGE. To determine protein levels in 3<sup>rd</sup> instar larva, 10 larvae were inverted in PBS, the gut, fat body and salivary glands were removed and the sample transferred into 100  $\mu$ l 2 x Laemmli buffer. After sonication for 10 sec in a Branson 250 sonicator at 30 pulse, intensity 3, the sample was boiled for 5 min at 95°C. The sonication was repeated and the sample subsequently cleared by 10 min high speed centrifugation. 10  $\mu$ l of the sample were used for PAGE. After wet transfer onto a PVDF membrane (Millipore) and 1 h incubation in blocking solution (5% milk, 0.3% Tween-20, PBS), membranes were probed for 1 h at RT with blocking solution containing 1:3.000 rabbit anti-MOF; 1:3.000 rabbit anti-MSL1; 1:1.000 rat anti-MSL3; 1:1.000 mouse anti-HA; 1:10.000 rabbit anti-H3; 1:1.000 rabbit anti H4; 1:1.000 rabbit anti-H4K16ac (Santa Cruz sc-8662-R); or rabbit anti-H4K5ac antibodies, respectively. Membranes were washed 3 x 5 min in PBS-T (0.3% Tween-20 in PBS) and incubated for 45 min with blocking solution containing 1:15.000 anti-mouse or anti-rabbit antibodies coupled

to horseradish peroxidase. After another 3 x 5 min wash with PBS-T, membranes were incubated for 5 min with Lumi-Light western blotting substrate (Roche). The luminescence signal was captured on Kodak Biomax MR films.

### **7.6 Expression and purification of *Drosophila* recombinant proteins with the baculovirus system**

The MOF variant proteins expressed in the baculovirus system were cloned as full length constructs in HA-pFastBac vectors. Recombinant baculoviruses were generated as described in the manual “Bac-to-Bac Baculovirus Expression System” (Invitrogen). Recombinant viruses for expression of MSL1 and FLAG-MSL3 were readily available in the lab. Recombinant viruses were used to infect insect cells (SF21) and these were harvested 2 days after infection by dissolving the cell pellets in HEMG-200 Buffer (25mM Hepes pH 7,6 ; 1mM EDTA ; 12,5 mM MgCl<sub>2</sub> ; 10 % Glycerol ; 200 mM KCl, 0,5% Triton X-100 reduced ; Complete Protease Inhibitor(Roche) ; 0,2 µM PMSF). For purification of the recombinant proteins, whole-cell-extracts were incubated for 2 h with HA-Agarose or Flag-Agarose-Beads (Sigma), followed by washes with HEMG-500 and HEMG-200. HA-and Flag-peptides (Sigma) were used in a concentration of 400 ng/µl to elute the recombinant proteins. All recombinant proteins were stored in HEMG 200 Buffer.

### **7.7 HAT-assay on nucleosomal templates**

HAT assays were performed by Herbert Holz. In a 20 µl reaction containing 50 mM Tris pH 8,0, 0,1 mM EDTA pH 8, 0, 1 mM DTT, 1 mM PMSF, 5% Glycerol, 10 mM Na-butyrate and 0,02 µCi of acetyl coenzyme A (acetyl-1-14C, 60 mCi/mmol (Perkin Elmer)) we incubated 5.3 nM (10 ng) of the respective recombinant MOF-protein (or trimeric complexes) with 1,5 µg of endogenous mono, -di- nucleosomes obtained from MCF-7 cells (human breast cancer cell line). For preparation of MCF-7 nucleosomes: See Epigenome Network of Excellence (NoE) protocols (<http://www.epigenome-noe.net/researchtools/protocol.php?protid=22>). After a 60 min incubation at 26°C the whole reaction volume was applied on a precast 12% BIS/Tris Novex Gel (Invitrogen) and run in 1x MES Buffer at 130V for 45 min and 155 V for another 45 min. Proteins were visualized by Coomassie R 250 Blue staining, destained, and the gel was dried on a 3 MM paper using a vacuum dryer (Biorad) at 80°C for 2 h. The dried gel was subsequently exposed on an imaging plate BAS-IP MS 2025 for 5 days. Acetylation signals were obtained by scanning the IP-MS plate on a FLA5000 scanner (Fujifilm).

### 7.8 Filter binding HAT-assay

Filter binding assays were performed by Herbert Holz. In a 30 µl reaction containing 50 mM Tris pH 8,0, 0,1 mM EDTA pH 8, 0, 1 mM DTT, 1 mM PMSF, 5% Glycerol, 10 mM Na-Butyrate and 0,02 µCi of acetyl coenzyme A (acetyl-1-14C, 60 mCi/mmol (Perkin Elmer)) we incubated 21.3 nM (40 ng) of the respective recombinant MOF-protein (or trimeric complexes) with 1,5 µg of recombinant (*Xenopus*) histone octamer. After a 60 min incubation at 26°C the whole reaction volume was applied on a P81 filter paper (Whatman, 1,5 x 1,5 cm), air dried and washed 3x10 min at RT in a large volume of 50 mM Na-carbonate pH 9,2. After rinsing the filters in acetone, they were air dried and counted in 5 ml of scintillation liquid (Rotiszint eco plus, Roth).

### 7.9 Electrophoretic mobility shift assay (EMSA)

The EMSA assay was performed by Ibrahim Ilik. N-terminal-6-His-tagged MOF chromobarrel domain constructs, including the wild-type chromobarrel domain (amino acids 346-448 of full-length MOF) and two point mutants Y416D and W426 were expressed in BL21 (DE3) cells and purified using Ni-NTA chromatography (Qiagen). The following template is used to express a 84nt radiolabeled ( $\alpha^{32}\text{P}$ -CTP) ssRNA using Riboprobe T7 kit (Promega):

CGTAATACGACTCACTATAGGGAGAAAGAACGAATATATATACGCGCAAATTAA  
GCAAATATATATGCATATATGGGAACGCGATTTTAATGAAGAGCTCTTC. The final RNA product of the *in vitro* transcription is underlined. EMSA was carried out with 25-100 ng of protein and 1µL of RNA (20kcpm/µL) in 20µL RNA binding buffer (20mM HEPES, NaOH (pH 7.6), 3mM MgCl<sub>2</sub>, 10% Glycerol, 1mM DTT, 0.1mg/ml BSA). The mixture was incubated on ice for 30 minutes and run on a 6% native polyacrylamide gel for 90 minutes at 120V. In order to visualize RNA-protein complexes, the gel was dried and exposed to an autoradiography film (Kodak XAR).

### 7.10 Immunoprecipitation of MOF constructs from SL-2 stable cell lines

Generation of stable cell lines and immunoprecipitations were performed by Erinc Hallacli. All MOF constructs used for SL-2 stable cell line generation carry N-terminal 3xFlag 6His tags and are under the control of an MtnA promoter. The vector also contains a Neomycin cassette, allowing omitting of the co-transfection step. 1 million cells were transfected with 0.5 µg of DNA with Qiagen Effectene reagent. The cells were selected initially with 1 mg/ml

Geneticin. During the amplification period, the concentration of Geneticin was gradually reduced to its final concentration of 0.25 mg/ml. For immunoprecipitation experiments, equal number of cells from different lines (FL MOF, MOF  $\Delta$ N, MOF Y416D) along with WT MOCK cells were induced with 100 mM CuSO<sub>4</sub> for one day and harvested. Nuclei were isolated and nuclear extracts were prepared by freeze-thaw cycles in HEMGT 150 buffer (25 mM HEPES 7.6, 0.1 mM EDTA, 12.5 mM MgCl<sub>2</sub>, 10% Glycerol, 0.2% Tween-20 and 150 mM KCl). INPUT samples were taken from the nuclear extract. Nuclear extracts were incubated with 30 ml bed volume of magnetic Flag Agarose beads (Sigma) for 2 hours at 4°C. The beads were washed four times with HEMGT 150 buffer and eluted with 3xFlag peptide (final 0.4 mg/ml concentration) containing HEMGT 150 buffer overnight. The elutions were precipitated by TCA/Acetone method and resuspended with 4 x Laemmli Buffer.

### 7.11 ChIP-Seq data analysis

All ChIP-Seq data analysis was performed by Florence Cavalli and Juanma Vaquerizas at the European Bioinformatics Institute.

#### 7.11.1 Data processing

ChIP samples were sequenced using Illumina GAIIX machines at the EMBL GeneCore facility. Resulting reads were mapped to the *D. melanogaster* genome (dmel\_r5.11\_FB2008\_08) using the Bowtie software [253] with the following parameters: -n 2 -k 1 --solexa1.3-quals --best. With these settings, the software maps reads only to unique locations in the genome, choosing the best possible hit. The software takes into consideration the read-quality during alignment and removes all reads containing ambiguous nucleotide assignments (N). Reads mapping to heterochromatic regions were discarded from further analysis (Table S1). We divided the *D. melanogaster* genome into non-overlapping 25 bp bins, and counted the number of reads mapped to each bin. These read-counts were then inputted to the DESeq BioConductor package [254]; counts were normalized between IP and input samples by applying a scaling factor accounting for differences in the total numbers of reads per sequencing run. Any bins with read counts of zero were discarded from further analysis. For each bin, DESeq outputted a log<sub>2</sub> fold-change (log<sub>2</sub>FC) value between normalized read counts in the IP and control samples. We used input DNA as the control for MOF-binding and histone H4-binding as the control for the H4K16 acetylation marks. As DNA fragment sizes after sonication was ~200 bp, we smoothed log<sub>2</sub>FC values using a 400 bp sliding window approach [215]. The final log<sub>2</sub>FC values represented the signal from the IP

samples relative to controls, with positive values corresponding to enrichments in the IP sample and negative values corresponding to enrichments in control sample. We calculated log2FC thresholds to indicate significant binding or acetylation compared with the control. Since negative log2FC values (i.e. enrichment in control signal) correspond to experimental noise, we fitted a symmetric null-distribution to the density distribution for log2FC values below the mode (PMID: 16732288). We then applied an FDR-adjusted p-value cut-off of 0.05 to identify 25 bp bins containing significant binding or acetylation. All ChIP-Seq data will be made available at European Nucleotide Archive (ENA) (<http://www.ebi.ac.uk/ena/>).

### **7.11.2 Detection of MOF-bound and H4K16-acetylated genes**

For the classification of binding patterns, gene bodies were defined as all exonic sequences between +500 bp downstream of the transcription start site (TSS) to the 3'-end, as annotated in the Ensembl database. The promoter was defined as the region between -200 bp and the TSS: the region was selected by identifying the mode for average MOF-binding at the 5'-end of all genes which falls about -100 bp upstream of the TSS in both male and female samples and then providing a 100 bp window either side. We excluded any genes that are <600 bp (1,205 loci) in length, as they were too short for this analysis. There is a clear bi-modal distribution dividing partially bound genes from fully bound ones. We used a threshold of 70% (relative to the gene body length) to differentiate between the two types of binding. Next, we classified partially bound genes as promoter bound, if they contained at least one bin with significant binding in the promoter region defined above. We used a similar classification to identify patterns of H4K16 acetylation. In this case, the promoter was defined as the region between the TSS and +500bp downstream, in order to accommodate shift in acetylation patterns towards the interior of genes compared with MOF-binding.

### **7.11.3 Gene expression profiling**

Gene expression was measured using Affymetrix Drosophila2 GeneChips in at least triplicate for wt male and female 3rd instar larva salivary glands. Data analysis was performed using publicly available packages in the BioConductor Software Suite [PMID: 15461798]. Raw .CEL files were processed using GCRMA and probe sets were mapped to genes using annotation available from the Ensembl database (v57) [PMID: 21045057]. Expressed genes were identified as those outputting MAS5.0 'present' calls in all available biological replicates. Microarray data will be available at ArrayExpress upon acceptance of the manuscript [PMID: 21071405].





# References

1. Finch, J.T., et al., *Structure of nucleosome core particles of chromatin*. Nature, 1977. **269**(5623): p. 29-36.
2. Klug, A., et al., *A low resolution structure for the histone core of the nucleosome*. Nature, 1980. **287**(5782): p. 509-16.
3. Talbert, P.B. and S. Henikoff, *Histone variants--ancient wrap artists of the epigenome*. Nat Rev Mol Cell Biol, 2010. **11**(4): p. 264-75.
4. Spadafora, C., et al., *The DNA repeat lengths in chromatins from sea urchin sperm and gastrule cells are markedly different*. FEBS Lett, 1976. **69**(1): p. 281-5.
5. Noll, M. and R.D. Kornberg, *Action of micrococcal nuclease on chromatin and the location of histone H1*. J Mol Biol, 1977. **109**(3): p. 393-404.
6. Ramakrishnan, V., *Histone H1 and chromatin higher-order structure*. Crit Rev Eukaryot Gene Expr, 1997. **7**(3): p. 215-30.
7. Bednar, J., et al., *Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin*. Proc Natl Acad Sci U S A, 1998. **95**(24): p. 14173-8.
8. Fan, Y., et al., *H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo*. Mol Cell Biol, 2003. **23**(13): p. 4559-72.
9. Maresca, T.J. and R. Heald, *The long and the short of it: linker histone H1 is required for metaphase chromosome compaction*. Cell Cycle, 2006. **5**(6): p. 589-91.
10. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 Å resolution*. Nature, 1997. **389**(6648): p. 251-60.
11. Schalch, T., et al., *X-ray structure of a tetranucleosome and its implications for the chromatin fibre*. Nature, 2005. **436**(7047): p. 138-41.
12. Woodcock, C.L. and R.P. Ghosh, *Chromatin higher-order structure and dynamics*. Cold Spring Harb Perspect Biol, 2010. **2**(5): p. a000596.
13. Eltsov, M., et al., *Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ*. Proc Natl Acad Sci U S A, 2008. **105**(50): p. 19732-7.
14. Clark, D.J. and T. Kimura, *Electrostatic mechanism of chromatin folding*. J Mol Biol, 1990. **211**(4): p. 883-96.
15. Widom, J. and A. Klug, *Structure of the 300Å chromatin filament: X-ray diffraction from oriented samples*. Cell, 1985. **43**(1): p. 207-13.
16. Woodcock, C.L. and J. Frank, *Nucleosome mass distribution using image averaging*. J Ultrastruct Res, 1984. **89**(3): p. 295-302.
17. Dorigo, B., et al., *Nucleosome arrays reveal the two-start organization of the chromatin fiber*. Science, 2004. **306**(5701): p. 1571-3.
18. Grigoryev, S.A., et al., *Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions*. Proc Natl Acad Sci U S A, 2009. **106**(32): p. 13317-22.
19. Muller, W.G., et al., *Generic features of tertiary chromatin structure as detected in natural chromosomes*. Mol Cell Biol, 2004. **24**(21): p. 9359-70.
20. Li, B., M. Carey, and J.L. Workman, *The role of chromatin during transcription*. Cell, 2007. **128**(4): p. 707-19.
21. Kireeva, M.L., et al., *Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription*. Mol Cell, 2002. **9**(3): p. 541-52.
22. Lacy, E. and R. Axel, *Analysis of DNA of isolated chromatin subunits*. Proc Natl Acad Sci U S A, 1975. **72**(10): p. 3978-82.
23. Lee, C.K., et al., *Evidence for nucleosome depletion at active regulatory regions genome-wide*. Nat Genet, 2004. **36**(8): p. 900-5.
24. Ray-Gallet, D. and G. Almouzni, *Nucleosome dynamics and histone variants*. Essays Biochem, 2010. **48**(1): p. 75-87.
25. De Koning, L., et al., *Histone chaperones: an escort network regulating histone traffic*. Nat Struct Mol Biol, 2007. **14**(11): p. 997-1007.

26. Polo, S.E. and G. Almouzni, *DNA damage leaves its mark on chromatin*. *Cell Cycle*, 2007. **6**(19): p. 2355-2359.
27. Zlatanova, J., C. Seebart, and M. Tomschik, *Nap1: taking a closer look at a juggler protein of extraordinary skills*. *FASEB J*, 2007. **21**(7): p. 1294-310.
28. Natsume, R., et al., *Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4*. *Nature*, 2007. **446**(7133): p. 338-41.
29. Schwabish, M.A. and K. Struhl, *Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II*. *Mol Cell*, 2006. **22**(3): p. 415-22.
30. Thiriet, C. and J.J. Hayes, *Replication-independent core histone dynamics at transcriptionally active loci in vivo*. *Genes Dev*, 2005. **19**(6): p. 677-82.
31. Kaplan, C.D., L. Laprade, and F. Winston, *Transcription elongation factors repress transcription initiation from cryptic sites*. *Science*, 2003. **301**(5636): p. 1096-9.
32. Tsubota, T., et al., *Histone H3-K56 acetylation is catalyzed by histone chaperone-dependent complexes*. *Mol Cell*, 2007. **25**(5): p. 703-12.
33. Downs, J.A., et al., *Binding of chromatin-modifying activities to phosphorylated histone H2A at DNA damage sites*. *Mol Cell*, 2004. **16**(6): p. 979-90.
34. Kaplan, N., et al., *The DNA-encoded nucleosome organization of a eukaryotic genome*. *Nature*, 2009. **458**(7236): p. 362-6.
35. Zhang, Y., et al., *Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo*. *Nat Struct Mol Biol*, 2009. **16**(8): p. 847-52.
36. Gilchrist, D.A., et al., *Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation*. *Cell*, 2010. **143**(4): p. 540-51.
37. Clapier, C.R. and B.R. Cairns, *The biology of chromatin remodeling complexes*. *Annu Rev Biochem*, 2009. **78**: p. 273-304.
38. Ferreira, H., A. Flaus, and T. Owen-Hughes, *Histone modifications influence the action of Snf2 family remodelling enzymes by different mechanisms*. *J Mol Biol*, 2007. **374**(3): p. 563-79.
39. Carey, M., B. Li, and J.L. Workman, *RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation*. *Mol Cell*, 2006. **24**(3): p. 481-7.
40. Armstrong, J.A., et al., *The Drosophila BRM complex facilitates global transcription by RNA polymerase II*. *EMBO J*, 2002. **21**(19): p. 5245-54.
41. Parnell, T.J., J.T. Huff, and B.R. Cairns, *RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes*. *EMBO J*, 2008. **27**(1): p. 100-10.
42. Boyer, L.A., R.R. Latek, and C.L. Peterson, *The SANT domain: a unique histone-tail-binding module?* *Nat Rev Mol Cell Biol*, 2004. **5**(2): p. 158-63.
43. Whitehouse, I., et al., *Chromatin remodelling at promoters suppresses antisense transcription*. *Nature*, 2007. **450**(7172): p. 1031-5.
44. Goldmark, J.P., et al., *The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p*. *Cell*, 2000. **103**(3): p. 423-33.
45. Badenhorst, P., et al., *Biological functions of the ISWI chromatin remodeling complex NURF*. *Genes Dev*, 2002. **16**(24): p. 3186-98.
46. Deuring, R., et al., *The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo*. *Mol Cell*, 2000. **5**(2): p. 355-65.
47. Flanagan, J.F., et al., *Double chromodomains cooperate to recognize the methylated histone H3 tail*. *Nature*, 2005. **438**(7071): p. 1181-5.
48. Sims, R.J., 3rd, et al., *Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains*. *J Biol Chem*, 2005. **280**(51): p. 41789-92.
49. Walfridsson, J., et al., *A genome-wide role for CHD remodelling factors and Nap1 in nucleosome disassembly*. *EMBO J*, 2007. **26**(12): p. 2868-79.
50. Marfella, C.G. and A.N. Imbalzano, *The Chd family of chromatin remodelers*. *Mutat Res*, 2007. **618**(1-2): p. 30-40.

51. Murawska, M., et al., *dCHD3, a novel ATP-dependent chromatin remodeler associated with sites of active transcription*. Mol Cell Biol, 2008. **28**(8): p. 2745-57.
52. Denslow, S.A. and P.A. Wade, *The human Mi-2/NuRD complex and gene regulation*. Oncogene, 2007. **26**(37): p. 5433-8.
53. Cai, Y., et al., *YY1 functions with INO80 to activate transcription*. Nat Struct Mol Biol, 2007. **14**(9): p. 872-4.
54. Hartley, P.D. and H.D. Madhani, *Mechanisms that specify promoter nucleosome location and identity*. Cell, 2009. **137**(3): p. 445-58.
55. Papamichos-Chronakis, M. and C.L. Peterson, *The Ino80 chromatin-remodeling enzyme regulates replisome function and stability*. Nat Struct Mol Biol, 2008. **15**(4): p. 338-45.
56. Shimada, K., et al., *Ino80 chromatin remodeling complex promotes recovery of stalled replication forks*. Curr Biol, 2008. **18**(8): p. 566-75.
57. van Attikum, H., et al., *Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand break repair*. Cell, 2004. **119**(6): p. 777-88.
58. Morrison, A.J. and X. Shen, *Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes*. Nat Rev Mol Cell Biol, 2009. **10**(6): p. 373-84.
59. Churchill, M.E. and M. Suzuki, *'SPKK' motifs prefer to bind to DNA at A/T-rich sites*. EMBO J, 1989. **8**(13): p. 4189-95.
60. Furuyama, T. and S. Henikoff, *Centromeric nucleosomes induce positive DNA supercoils*. Cell, 2009. **138**(1): p. 104-13.
61. Dalal, Y., et al., *Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells*. PLoS Biol, 2007. **5**(8): p. e218.
62. Schwartz, B.E. and K. Ahmad, *Transcriptional activation triggers deposition and removal of the histone variant H3.3*. Genes Dev, 2005. **19**(7): p. 804-14.
63. Sutcliffe, E.L., et al., *Dynamic histone variant exchange accompanies gene induction in T cells*. Mol Cell Biol, 2009. **29**(7): p. 1972-86.
64. Ahmad, K. and S. Henikoff, *The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly*. Mol Cell, 2002. **9**(6): p. 1191-200.
65. Mito, Y., J.G. Henikoff, and S. Henikoff, *Genome-scale profiling of histone H3.3 replacement patterns*. Nat Genet, 2005. **37**(10): p. 1090-7.
66. Ng, R.K. and J.B. Gurdon, *Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription*. Nat Cell Biol, 2008. **10**(1): p. 102-9.
67. Ishibashi, T., et al., *Acetylation of vertebrate H2A.Z and its effect on the structure of the nucleosome*. Biochemistry, 2009. **48**(22): p. 5007-17.
68. Suto, R.K., et al., *Crystal structure of a nucleosome core particle containing the variant histone H2A.Z*. Nat Struct Biol, 2000. **7**(12): p. 1121-4.
69. Dorigo, B., et al., *Chromatin fiber folding: requirement for the histone H4 N-terminal tail*. J Mol Biol, 2003. **327**(1): p. 85-96.
70. Adam, M., et al., *H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions*. Mol Cell Biol, 2001. **21**(18): p. 6270-9.
71. Hardy, S., et al., *The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition*. PLoS Genet, 2009. **5**(10): p. e1000687.
72. Zlatanova, J. and A. Thakar, *H2A.Z: view from the top*. Structure, 2008. **16**(2): p. 166-79.
73. Altaf, M., et al., *Connection between histone H2A variants and chromatin remodeling complexes*. Biochem Cell Biol, 2009. **87**(1): p. 35-50.
74. Brickner, D.G., et al., *H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state*. PLoS Biol, 2007. **5**(4): p. e81.
75. van Attikum, H. and S.M. Gasser, *Crosstalk between histone modifications during the DNA damage response*. Trends Cell Biol, 2009. **19**(5): p. 207-17.

76. Fernandez-Capetillo, O., et al., *H2AX is required for chromatin remodeling and inactivation of sex chromosomes in male mouse meiosis*. Dev Cell, 2003. **4**(4): p. 497-508.
77. Turner, J.M., et al., *Silencing of unsynapsed meiotic chromosomes in the mouse*. Nat Genet, 2005. **37**(1): p. 41-7.
78. Doyen, C.M., et al., *Mechanism of polymerase II transcription repression by the histone variant macroH2A*. Mol Cell Biol, 2006. **26**(3): p. 1156-64.
79. Buschbeck, M., et al., *The histone variant macroH2A is an epigenetic regulator of key developmental genes*. Nat Struct Mol Biol, 2009. **16**(10): p. 1074-9.
80. Chadwick, B.P. and H.F. Willard, *Cell cycle-dependent localization of macroH2A in chromatin of the inactive X chromosome*. J Cell Biol, 2002. **157**(7): p. 1113-23.
81. Timinszky, G., et al., *A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation*. Nat Struct Mol Biol, 2009. **16**(9): p. 923-9.
82. Allfrey, V.G., R. Faulkner, and A.E. Mirsky, *Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis*. Proc Natl Acad Sci U S A, 1964. **51**: p. 786-94.
83. Brownell, J.E., et al., *Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation*. Cell, 1996. **84**(6): p. 843-51.
84. Bannister, A.J. and T. Kouzarides, *The CBP co-activator is a histone acetyltransferase*. Nature, 1996. **384**(6610): p. 641-3.
85. Chen, H., et al., *Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300*. Cell, 1997. **90**(3): p. 569-80.
86. Ogryzko, V.V., et al., *The transcriptional coactivators p300 and CBP are histone acetyltransferases*. Cell, 1996. **87**(5): p. 953-9.
87. Spencer, T.E., et al., *Steroid receptor coactivator-1 is a histone acetyltransferase*. Nature, 1997. **389**(6647): p. 194-8.
88. Taunton, J., C.A. Hassig, and S.L. Schreiber, *A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p*. Science, 1996. **272**(5260): p. 408-11.
89. Kouzarides, T., *Chromatin modifications and their function*. Cell, 2007. **128**(4): p. 693-705.
90. Smith, E. and A. Shilatifard, *The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes*. Mol Cell, 2010. **40**(5): p. 689-701.
91. Bhaumik, S.R., E. Smith, and A. Shilatifard, *Covalent modifications of histones during development and disease pathogenesis*. Nat Struct Mol Biol, 2007. **14**(11): p. 1008-16.
92. Brower-Toland, B., et al., *Specific contributions of histone tails and their acetylation to the mechanical stability of nucleosomes*. J Mol Biol, 2005. **346**(1): p. 135-46.
93. Dunker, A.K., et al., *Intrinsically disordered protein*. J Mol Graph Model, 2001. **19**(1): p. 26-59.
94. Anderson, J.D. and J. Widom, *Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites*. Mol Cell Biol, 2001. **21**(11): p. 3830-9.
95. Gorisch, S.M., et al., *Histone acetylation increases chromatin accessibility*. J Cell Sci, 2005. **118**(Pt 24): p. 5825-34.
96. Morales, V. and H. Richard-Foy, *Role of histone N-terminal tails and their acetylation in nucleosome dynamics*. Mol Cell Biol, 2000. **20**(19): p. 7230-7.
97. Protacio, R.U., et al., *Effects of histone tail domains on the rate of transcriptional elongation through a nucleosome*. Mol Cell Biol, 2000. **20**(23): p. 8866-78.
98. Shogren-Knaak, M., et al., *Histone H4-K16 acetylation controls chromatin structure and protein interactions*. Science, 2006. **311**(5762): p. 844-7.
99. Bannister, A.J., et al., *Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain*. Nature, 2001. **410**(6824): p. 120-4.
100. Lachner, M., et al., *Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins*. Nature, 2001. **410**(6824): p. 116-20.
101. Motamedi, M.R., et al., *HP1 proteins form distinct complexes and mediate heterochromatic gene silencing by nonoverlapping mechanisms*. Mol Cell, 2008. **32**(6): p. 778-90.

102. Rea, S., et al., *Regulation of chromatin structure by site-specific histone H3 methyltransferases*. Nature, 2000. **406**(6796): p. 593-9.
103. Stewart, M.D., J. Li, and J. Wong, *Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment*. Mol Cell Biol, 2005. **25**(7): p. 2525-38.
104. Fischle, W., et al., *Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation*. Nature, 2005. **438**(7071): p. 1116-22.
105. Kuzmichev, A., et al., *Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein*. Genes Dev, 2002. **16**(22): p. 2893-905.
106. Duncan, I.M., *Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of Drosophila melanogaster*. Genetics, 1982. **102**(1): p. 49-70.
107. Ringrose, L. and R. Paro, *Polycomb/Trithorax response elements and epigenetic memory of cell identity*. Development, 2007. **134**(2): p. 223-32.
108. Simon, J.A. and R.E. Kingston, *Mechanisms of polycomb gene silencing: knowns and unknowns*. Nat Rev Mol Cell Biol, 2009. **10**(10): p. 697-708.
109. Wood, A., et al., *The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p*. J Biol Chem, 2003. **278**(37): p. 34739-42.
110. Krogan, N.J., et al., *The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation*. Mol Cell, 2003. **11**(3): p. 721-9.
111. Ruthenburg, A.J., C.D. Allis, and J. Wysocka, *Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark*. Mol Cell, 2007. **25**(1): p. 15-30.
112. Vermeulen, M., et al., *Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4*. Cell, 2007. **131**(1): p. 58-69.
113. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications*. Nature, 2000. **403**(6765): p. 41-5.
114. Vakoc, C.R., et al., *Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin*. Mol Cell, 2005. **19**(3): p. 381-91.
115. Carrozza, M.J., et al., *Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription*. Cell, 2005. **123**(4): p. 581-92.
116. Joshi, A.A. and K. Struhl, *Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation*. Mol Cell, 2005. **20**(6): p. 971-8.
117. Keogh, M.C., et al., *Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex*. Cell, 2005. **123**(4): p. 593-605.
118. Eisen, A., et al., *The yeast NuA4 and Drosophila MSL complexes contain homologous subunits important for transcription regulation*. J Biol Chem, 2001. **276**(5): p. 3484-91.
119. Shi, X., et al., *ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression*. Nature, 2006. **442**(7098): p. 96-9.
120. Huang, Y., et al., *Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A*. Science, 2006. **312**(5774): p. 748-51.
121. Zippo, A., et al., *Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation*. Cell, 2009. **138**(6): p. 1122-36.
122. Zippo, A., et al., *PIM1-dependent phosphorylation of histone H3 at serine 10 is required for MYC-dependent transcriptional activation and oncogenic transformation*. Nat Cell Biol, 2007. **9**(8): p. 932-44.
123. Macdonald, N., et al., *Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14-3-3*. Mol Cell, 2005. **20**(2): p. 199-211.
124. Soloaga, A., et al., *MSK2 and MSK1 mediate the mitogen- and stress-induced phosphorylation of histone H3 and HMG-14*. EMBO J, 2003. **22**(11): p. 2788-97.

125. Wang, Y., et al., *The JIL-1 tandem kinase mediates histone H3 phosphorylation and is required for maintenance of chromatin structure in Drosophila*. Cell, 2001. **105**(4): p. 433-43.
126. Prestel, M., C. Feller, and P.B. Becker, *Dosage compensation and the global re-balancing of aneuploid genomes*. Genome Biol, 2010. **11**(8): p. 216.
127. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans*. Nature, 1998. **391**(6669): p. 806-11.
128. van Wolfswinkel, J.C. and R.F. Ketting, *The role of small non-coding RNAs in genome stability and chromatin organization*. J Cell Sci, 2010. **123**(Pt 11): p. 1825-39.
129. Chueh, A.C., et al., *LINE retrotransposon RNA is an essential structural and functional epigenetic component of a core neocentromeric chromatin*. PLoS Genet, 2009. **5**(1): p. e1000354.
130. Wong, L.H., et al., *Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere*. Genome Res, 2007. **17**(8): p. 1146-60.
131. Fukagawa, T., et al., *Dicer is essential for formation of the heterochromatin structure in vertebrate cells*. Nat Cell Biol, 2004. **6**(8): p. 784-91.
132. Czech, B., et al., *An endogenous small interfering RNA pathway in Drosophila*. Nature, 2008. **453**(7196): p. 798-802.
133. Ghildiyal, M., et al., *Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells*. Science, 2008. **320**(5879): p. 1077-81.
134. Kawamura, Y., et al., *Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells*. Nature, 2008. **453**(7196): p. 793-7.
135. Saito, K. and M.C. Siomi, *Small RNA-mediated quiescence of transposable elements in animals*. Dev Cell, 2010. **19**(5): p. 687-97.
136. Kanhere, A., et al., *Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2*. Mol Cell, 2010. **38**(5): p. 675-88.
137. Mercer, T.R., M.E. Dinger, and J.S. Mattick, *Long non-coding RNAs: insights into functions*. Nat Rev Genet, 2009. **10**(3): p. 155-9.
138. Struhl, K., *Transcriptional noise and the fidelity of initiation by RNA polymerase II*. Nat Struct Mol Biol, 2007. **14**(2): p. 103-5.
139. Ponjavic, J., C.P. Ponting, and G. Lunter, *Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs*. Genome Res, 2007. **17**(5): p. 556-65.
140. Amaral, P.P. and J.S. Mattick, *Noncoding RNA in development*. Mamm Genome, 2008. **19**(7-8): p. 454-92.
141. Roy, S., et al., *Identification of functional elements and regulatory circuits by Drosophila modENCODE*. Science, 2010. **330**(6012): p. 1787-97.
142. Gerstein, M.B., et al., *Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project*. Science, 2010. **330**(6012): p. 1775-87.
143. Penny, G.D., et al., *Requirement for Xist in X chromosome inactivation*. Nature, 1996. **379**(6561): p. 131-7.
144. Zhao, J., et al., *Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome*. Science, 2008. **322**(5902): p. 750-6.
145. Chow, J. and E. Heard, *X inactivation and the complexities of silencing a sex chromosome*. Curr Opin Cell Biol, 2009. **21**(3): p. 359-66.
146. Nagano, T., et al., *The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin*. Science, 2008. **322**(5908): p. 1717-20.
147. Pandey, R.R., et al., *Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation*. Mol Cell, 2008. **32**(2): p. 232-46.
148. Tsai, M.C., et al., *Long noncoding RNA as modular scaffold of histone modification complexes*. Science, 2010. **329**(5992): p. 689-93.
149. Huarte, M., et al., *A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response*. Cell, 2010. **142**(3): p. 409-19.

150. Zhao, J., et al., *Genome-wide identification of polycomb-associated RNAs by RIP-seq*. Mol Cell, 2010. **40**(6): p. 939-53.
151. Wang, X., et al., *Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription*. Nature, 2008. **454**(7200): p. 126-30.
152. Feng, J., et al., *The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator*. Genes Dev, 2006. **20**(11): p. 1470-84.
153. Ohno, M., et al., *Triplex-forming DNAs in the human interphase nucleus visualized in situ by polypurine/polypyrimidine DNA probes and antitriplex antibodies*. Chromosoma, 2002. **111**(3): p. 201-13.
154. Mariner, P.D., et al., *Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock*. Mol Cell, 2008. **29**(4): p. 499-509.
155. Beltran, M., et al., *A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition*. Genes Dev, 2008. **22**(6): p. 756-69.
156. Chen, L.L. and G.G. Carmichael, *Decoding the function of nuclear long non-coding RNAs*. Curr Opin Cell Biol, 2010. **22**(3): p. 357-64.
157. Heitz, E., *Das Heterochromatin der Moose*. Jahrb Wiss Bot, 1928. **69**(1): p. 762-818.
158. Gilbert, N., et al., *Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers*. Cell, 2004. **118**(5): p. 555-66.
159. Yasuhara, J.C. and B.T. Wakimoto, *Oxymoron no more: the expanding world of heterochromatic genes*. Trends Genet, 2006. **22**(6): p. 330-8.
160. Schotta, G., et al., *Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila*. Semin Cell Dev Biol, 2003. **14**(1): p. 67-75.
161. Kharchenko, P.V., et al., *Comprehensive analysis of the chromatin landscape in Drosophila melanogaster*. Nature, 2010.
162. Fillion, G.J., et al., *Systematic protein location mapping reveals five principal chromatin types in Drosophila cells*. Cell, 2010. **143**(2): p. 212-24.
163. Lindsley, D.L., et al., *Segmental aneuploidy and the genetic gross structure of the Drosophila genome*. Genetics, 1972. **71**(1): p. 157-84.
164. Wilson, M.A. and K.D. Makova, *Genomic analyses of sex chromosome evolution*. Annu Rev Genomics Hum Genet, 2009. **10**: p. 333-54.
165. Kaiser, V.B. and D. Bachtrog, *Evolution of sex chromosomes in insects*. Annu Rev Genet, 2010. **44**: p. 91-112.
166. Georgiev, P., S. Chlamydas, and A. Akhtar, *Drosophila dosage compensation: Males are from Mars, females are from Venus*. Fly (Austin), 2011. **5**(2).
167. Chow, J.C. and E. Heard, *Nuclear organization and dosage compensation*. Cold Spring Harb Perspect Biol, 2010. **2**(11): p. a000604.
168. Meyer, B.J., *Targeting X chromosomes for repression*. Curr Opin Genet Dev, 2010. **20**(2): p. 179-89.
169. Gupta, V., et al., *Global analysis of X-chromosome dosage compensation*. J Biol, 2006. **5**(1): p. 3.
170. Nguyen, D.K. and C.M. Disteché, *Dosage compensation of the active X chromosome in mammals*. Nat Genet, 2006. **38**(1): p. 47-53.
171. Wolf, J.B. and J. Bryk, *General lack of global dosage compensation in ZZ/ZW systems? Broadening the perspective with RNA-seq*. BMC Genomics, 2011. **12**: p. 91.
172. Gelbart, M.E. and M.I. Kuroda, *Drosophila dosage compensation: a complex voyage to the X chromosome*. Development, 2009. **136**(9): p. 1399-410.
173. Bridges, C.B., *Sex in relation to chromosomes and genes*. Am. Nat., 1922(56): p. 51-63.
174. Muller, H.J., *Evidence of the precision of genetic adaptation*. The Harvey Lectures, 1950(18): p. 165-229.
175. Komma, D.J., *Effect of sex transformation genes on glucose-6-phosphate dehydrogenase activity in Drosophila melanogaster*. Genetics, 1966. **54**(2): p. 497-503.



176. Mukherjee, A.S. and W. Beermann, *Synthesis of ribonucleic acid by the X-chromosomes of Drosophila melanogaster and the problem of dosage compensation*. Nature, 1965. **207**(998): p. 785-6.
177. Fukunaga, A., A. Tanaka, and K. Oishi, *Maleless, a recessive autosomal mutant of Drosophila melanogaster that specifically kills male zygotes*. Genetics, 1975. **81**(1): p. 135-41.
178. Belote, J.M. and J.C. Lucchesi, *Male-specific lethal mutations of Drosophila melanogaster*. Genetics, 1980. **96**(1): p. 165-86.
179. Bone, J.R., et al., *Acetylated histone H4 on the male X chromosome is associated with dosage compensation in Drosophila*. Genes Dev, 1994. **8**(1): p. 96-104.
180. Gorman, M., A. Franke, and B.S. Baker, *Molecular characterization of the male-specific lethal-3 gene and investigations of the regulation of dosage compensation in Drosophila*. Development, 1995. **121**(2): p. 463-75.
181. Kelley, R.L., et al., *Expression of msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in Drosophila*. Cell, 1995. **81**(6): p. 867-77.
182. Kelley, R.L., et al., *Sex lethal controls dosage compensation in Drosophila by a non-splicing mechanism*. Nature, 1997. **387**(6629): p. 195-9.
183. Bashaw, G.J. and B.S. Baker, *The regulation of the Drosophila msl-2 gene reveals a function for Sex-lethal in translational control*. Cell, 1997. **89**(5): p. 789-98.
184. Turner, B.M., A.J. Birley, and J. Lavender, *Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei*. Cell, 1992. **69**(2): p. 375-84.
185. Gu, W., P. Szauter, and J.C. Lucchesi, *Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of Drosophila melanogaster*. Dev Genet, 1998. **22**(1): p. 56-64.
186. Hilfiker, A., et al., *mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in Drosophila*. EMBO J, 1997. **16**(8): p. 2054-60.
187. Akhtar, A. and P.B. Becker, *Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila*. Mol Cell, 2000. **5**(2): p. 367-75.
188. Copps, K., et al., *Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly*. EMBO J, 1998. **17**(18): p. 5409-17.
189. Amrein, H. and R. Axel, *Genes expressed in neurons of adult male Drosophila*. Cell, 1997. **88**(4): p. 459-69.
190. Meller, V.H., et al., *roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system*. Cell, 1997. **88**(4): p. 445-57.
191. Meller, V.H. and B.P. Rattner, *The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex*. EMBO J, 2002. **21**(5): p. 1084-91.
192. Franke, A. and B.S. Baker, *The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila*. Mol Cell, 1999. **4**(1): p. 117-22.
193. Lee, C.G. and J. Hurwitz, *Human RNA helicase A is homologous to the maleless protein of Drosophila*. J Biol Chem, 1993. **268**(22): p. 16822-30.
194. Gu, W., et al., *Targeting the chromatin-remodeling MSL complex of Drosophila to its sites of action on the X chromosome requires both acetyl transferase and ATPase activities*. EMBO J, 2000. **19**(19): p. 5202-11.
195. Richter, L., J.R. Bone, and M.I. Kuroda, *RNA-dependent association of the Drosophila maleless protein with the male X chromosome*. Genes Cells, 1996. **1**(3): p. 325-36.
196. Kelley, R.L., et al., *Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin*. Cell, 1999. **98**(4): p. 513-22.
197. Palmer, M.J., et al., *Sex-specific regulation of the male-specific lethal-1 dosage compensation gene in Drosophila*. Genes Dev, 1994. **8**(6): p. 698-706.

198. Dahlsveen, I.K., et al., *Targeting determinants of dosage compensation in Drosophila*. PLoS Genet, 2006. **2**(2): p. e5.
199. Alekseyenko, A.A., et al., *A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome*. Cell, 2008. **134**(4): p. 599-609.
200. Straub, T., et al., *The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex*. PLoS Genet, 2008. **4**(12): p. e1000302.
201. Alekseyenko, A.A., et al., *High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome*. Genes Dev, 2006. **20**(7): p. 848-57.
202. Gilfillan, G.D., et al., *Chromosome-wide gene-specific targeting of the Drosophila dosage compensation complex*. Genes Dev, 2006. **20**(7): p. 858-70.
203. Legube, G., et al., *X-chromosome-wide profiling of MSL-1 distribution and dosage compensation in Drosophila*. Genes Dev, 2006. **20**(7): p. 871-83.
204. Grimaud, C. and P.B. Becker, *The dosage compensation complex shapes the conformation of the X chromosome in Drosophila*. Genes Dev, 2009. **23**(21): p. 2490-5.
205. Li, F., A.H. Schiemann, and M.J. Scott, *Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the Drosophila MSL1/MSL2 complex*. Mol Cell Biol, 2008. **28**(4): p. 1252-64.
206. Grimaud, C. and P.B. Becker, *Form and function of dosage-compensated chromosomes--a chicken-and-egg relationship*. Bioessays, 2010. **32**(8): p. 709-17.
207. Demakova, O.V., et al., *The MSL complex levels are critical for its correct targeting to the chromosomes in Drosophila melanogaster*. Chromosoma, 2003. **112**(3): p. 103-15.
208. Kind, J. and A. Akhtar, *Cotranscriptional recruitment of the dosage compensation complex to X-linked target genes*. Genes Dev, 2007. **21**(16): p. 2030-40.
209. Gorchakov, A.A., et al., *Long-range spreading of dosage compensation in Drosophila captures transcribed autosomal genes inserted on X*. Genes Dev, 2009. **23**(19): p. 2266-71.
210. Larschan, E., et al., *MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism*. Mol Cell, 2007. **28**(1): p. 121-33.
211. Bell, O., et al., *Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in Drosophila melanogaster*. Mol Cell Biol, 2008. **28**(10): p. 3401-9.
212. Sural, T.H., et al., *The MSL3 chromodomain directs a key targeting step for dosage compensation of the Drosophila melanogaster X chromosome*. Nat Struct Mol Biol, 2008. **15**(12): p. 1318-25.
213. Kim, D., et al., *Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain*. Nat Struct Mol Biol, 2010. **17**(8): p. 1027-9.
214. Moore, S.A., et al., *Structural and biochemical studies on the chromo-barrel domain of male specific lethal 3 (MSL3) reveal a binding preference for mono- or dimethyllysine 20 on histone H4*. J Biol Chem, 2010. **285**(52): p. 40879-90.
215. Kind, J., et al., *Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila*. Cell, 2008. **133**(5): p. 813-28.
216. Vaquerizas, J.M., et al., *Nuclear pore proteins nup153 and megator define transcriptionally active regions in the Drosophila genome*. PLoS Genet, 2010. **6**(2): p. e1000846.
217. Larschan, E., et al., *X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila*. Nature, 2011. **471**(7336): p. 115-8.
218. Jin, Y., et al., *JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex*. J Cell Biol, 2000. **149**(5): p. 1005-10.
219. Regnard, C., et al., *Global Analysis of the Relationship between JIL-1 Kinase and Transcription*. PLoS Genet, 2011. **7**(3): p. e1001327.
220. Spierer, A., et al., *SU(VAR)3-7 links heterochromatin and dosage compensation in Drosophila*. PLoS Genet, 2008. **4**(5): p. e1000066.

221. Spierer, A., et al., *Loss of the modifiers of variegation Su(var)3-7 or HP1 impacts male X polytene chromosome morphology and dosage compensation*. J Cell Sci, 2005. **118**(Pt 21): p. 5047-57.
222. Furuhashi, H., M. Nakajima, and S. Hirose, *DNA supercoiling factor contributes to dosage compensation in Drosophila*. Development, 2006. **133**(22): p. 4475-83.
223. Patalano, S., et al., *Dual sex-specific functions of Drosophila Upstream of N-ras in the control of X chromosome dosage compensation*. Development, 2009. **136**(4): p. 689-98.
224. Mendjan, S., et al., *Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila*. Mol Cell, 2006. **21**(6): p. 811-23.
225. Stenberg, P., et al., *Buffering of segmental and chromosomal aneuploidies in Drosophila melanogaster*. PLoS Genet, 2009. **5**(5): p. e1000465.
226. Zhang, Y., et al., *Expression in aneuploid Drosophila S2 cells*. PLoS Biol, 2010. **8**(2): p. e1000320.
227. Raja, S.J., et al., *The nonspecific lethal complex is a transcriptional regulator in Drosophila*. Mol Cell, 2010. **38**(6): p. 827-41.
228. Bell, O., et al., *Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila*. EMBO J, 2007. **26**(24): p. 4974-84.
229. Schwaiger, M., et al., *Chromatin state marks cell-type- and gender-specific replication of the Drosophila genome*. Genes Dev, 2009. **23**(5): p. 589-601.
230. Gelbart, M.E., et al., *Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation*. Nat Struct Mol Biol, 2009. **16**(8): p. 825-32.
231. Morales, V., et al., *Functional integration of the histone acetyltransferase MOF into the dosage compensation complex*. EMBO J, 2004. **23**(11): p. 2258-68.
232. Kadlec, J., et al., *Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1*. Nat Struct Mol Biol, 2011. **18**(2): p. 142-9.
233. Akhtar, A. and P.B. Becker, *The histone H4 acetyltransferase MOF uses a C2HC zinc finger for substrate recognition*. EMBO Rep, 2001. **2**(2): p. 113-8.
234. Buscaino, A., G. Legube, and A. Akhtar, *X-chromosome targeting and dosage compensation are mediated by distinct domains in MSL-3*. EMBO Rep, 2006. **7**(5): p. 531-8.
235. Nielsen, P.R., et al., *Structure of the chromo barrel domain from the MOF acetyltransferase*. J Biol Chem, 2005. **280**(37): p. 32326-31.
236. Akhtar, A., D. Zink, and P.B. Becker, *Chromodomains are protein-RNA interaction modules*. Nature, 2000. **407**(6802): p. 405-9.
237. Sanjuan, R. and I. Marin, *Tracing the origin of the compensasome: evolutionary history of DEAH helicase and MYST acetyltransferase gene families*. Mol Biol Evol, 2001. **18**(3): p. 330-43.
238. Park, S.W., et al., *An evolutionarily conserved domain of roX2 RNA is sufficient for induction of H4-Lys16 acetylation on the Drosophila X chromosome*. Genetics, 2007. **177**(3): p. 1429-37.
239. Park, S.W., M.I. Kuroda, and Y. Park, *Regulation of histone H4 Lys16 acetylation by predicted alternative secondary structures in roX noncoding RNAs*. Mol Cell Biol, 2008. **28**(16): p. 4952-62.
240. Suganuma, T., et al., *ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding*. Nat Struct Mol Biol, 2008. **15**(4): p. 364-72.
241. Taipale, M., et al., *hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells*. Mol Cell Biol, 2005. **25**(15): p. 6798-810.
242. Smith, E.R., et al., *A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16*. Mol Cell Biol, 2005. **25**(21): p. 9175-88.
243. Wang, Z., et al., *Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes*. Cell, 2009. **138**(5): p. 1019-31.

- 
244. Kimura, A., T. Umehara, and M. Horikoshi, *Chromosomal gradient of histone acetylation established by Sas2p and Sir2p functions as a shield against gene silencing*. Nat Genet, 2002. **32**(3): p. 370-7.
245. Suka, N., K. Luo, and M. Grunstein, *Sir2p and Sas2p opposingly regulate acetylation of yeast histone H4 lysine16 and spreading of heterochromatin*. Nat Genet, 2002. **32**(3): p. 378-83.
246. Robert, F., et al., *Global position and recruitment of HATs and HDACs in the yeast genome*. Mol Cell, 2004. **16**(2): p. 199-209.
247. Bone, J.R. and M.I. Kuroda, *Dosage compensation regulatory proteins and the evolution of sex chromosomes in Drosophila*. Genetics, 1996. **144**(2): p. 705-13.
248. Bachtrög, D., *Positive selection at the binding sites of the male-specific lethal complex involved in dosage compensation in Drosophila*. Genetics, 2008. **180**(2): p. 1123-9.
249. Rodriguez, M.A., et al., *Species-specific positive selection of the male-specific lethal complex that participates in dosage compensation in Drosophila*. Proc Natl Acad Sci U S A, 2007. **104**(39): p. 15412-7.
250. Chouard, T., *Structural biology: Breaking the protein rules*. Nature, 2011. **471**(7337): p. 151-3.
251. Oldfield, C.J., et al., *Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners*. BMC Genomics, 2008. **9 Suppl 1**: p. S1.
252. Rea, S., G. Xouri, and A. Akhtar, *Males absent on the first (MOF): from flies to humans*. Oncogene, 2007. **26**(37): p. 5385-94.
253. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome*. Genome Biol, 2009. **10**(3): p. R25.
254. Anders, S. and W. Huber, *Differential expression analysis for sequence count data*. Genome Biol, 2010. **11**(10): p. R106.

## **Appendix A**

**“Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in *Drosophila melanogaster*.”**

Bell O, Conrad T, Kind J, Wirbelauer C, Akhtar A, Schübeler D.

Mol Cell Biol. 2008 May;28(10):3401-9. Epub 2008 Mar 17.



# Transcription-Coupled Methylation of Histone H3 at Lysine 36 Regulates Dosage Compensation by Enhancing Recruitment of the MSL Complex in *Drosophila melanogaster*<sup>∇</sup>

Oliver Bell,<sup>1</sup> Thomas Conrad,<sup>2,†</sup> Jop Kind,<sup>2,†</sup> Christiane Wirbelauer,<sup>1</sup>  
Asifa Akhtar,<sup>2,\*</sup> and Dirk Schübeler<sup>1,\*</sup>

Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland,<sup>1</sup> and  
European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany<sup>2</sup>

Received 3 January 2008/Returned for modification 28 January 2008/Accepted 10 March 2008

**In *Drosophila melanogaster*, dosage compensation relies on the targeting of the male-specific lethal (MSL) complex to hundreds of sites along the male X chromosome. Transcription-coupled methylation of histone H3 lysine 36 is enriched toward the 3' end of active genes, similar to the MSL proteins. Here, we have studied the link between histone H3 methylation and MSL complex targeting using RNA interference and chromatin immunoprecipitation. We show that trimethylation of histone H3 at lysine 36 (H3K36me3) relies on the histone methyltransferase Hypb and is localized promoter distal at dosage-compensated genes, similar to active genes on autosomes. However, H3K36me3 has an X-specific function, as reduction specifically decreases acetylation of histone H4 lysine 16 on the male X chromosome. This hypoacetylation is caused by compromised MSL binding and results in a failure to increase expression twofold. Thus, H3K36me3 marks the body of all active genes yet is utilized in a chromosome-specific manner to enhance histone acetylation at sites of dosage compensation.**

Similar to sex determination in mammals, sex determination in flies involves an unequal distribution of sex chromosomes, with females carrying two X chromosomes compared to one in males. The resulting difference in gene dose requires compensating mechanisms that guarantee equal expression of X-linked genes in both sexes. In mammals, dosage compensation involves transcriptional silencing of one of the two female copies of the X chromosome. In flies, the opposite strategy is realized, as transcription on the single male X chromosome is increased twofold.

Genetic screens in flies identified five proteins and two non-coding RNAs that are essential for the process of dosage compensation and whose absence causes male-specific lethality (MSL). The proteins MSL1, MSL2, MSL3, MLE (maleless), and MOF (males absent on the first) form a ribonucleoprotein complex (the MSL complex), together with the two noncoding RNAs, *roX1* and *roX2* (RNA on the X), which specifically binds to hundreds of sites on the male X chromosome and promotes transcriptional upregulation (16, 18, 25). This transcriptional upregulation is believed to involve histone hyperacetylation of dosage-compensated genes by the histone H4 lysine 16-specific histone acetyltransferase MOF (10).

Recent studies suggested that MSL complex binding sites fall into two categories: high-affinity sites which are able to

bind partial MSL complexes consisting of MSL1 and MSL2 (in mutant backgrounds of MSL3, MLE, and MOF) and low-affinity sites which require the full complement of the MSL complex (8, 12, 17). It has been postulated that the presence of high-affinity sites creates high local concentrations of MSL complex members on the X chromosome, which subsequently allows MSL complex binding to sites of lower affinity. Chromatin immunoprecipitation coupled with DNA microarrays (ChIP-chip) profiling experiments showed that the MSL complex localizes primarily to the 3' end of dosage-compensated genes (2, 9). Indeed, DNA elements in the 3' end of these target genes have been shown to be required for MSL binding, yet the ability to recruit the MSL complex strongly depends on their transcriptional activity (6, 14). This suggests that recognition of target DNA sequences with relatively low affinity for MSL proteins either is dependent on high chromatin accessibility or requires an additional, transcription-coupled signal.

Interestingly, trimethylation of histone H3 at lysine 36 (H3K36me3) is a histone modification that has been shown to be enriched specifically toward the 3' end of active genes (3, 20, 22, 23). In *Saccharomyces cerevisiae*, H3K36me is bound by the chromodomain protein Eaf3 and recruits the Rpd3S HDAC complex to remove transcription-coupled hyperacetylation, which could otherwise unmask internal transcription start sites (5, 11, 13). MSL3 is a *Drosophila* homologue of yeast Eaf3, opening the possibility that it interacts with methylated H3 lysine 36 to recruit the MSL complex to the 3' end of dosage-compensated genes.

In this study, we investigated the relationship between H3K36 methylation and MSL complex recruitment using RNA interference (RNAi) and ChIP in male *Drosophila* SL2 cells. We show that H3K36me3 is enriched promoter distal at dosage-compensated genes and relies on the histone methyltrans-

\* Corresponding author. Mailing address for Asifa Akhtar: European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. Phone: 49 6221 387 550. Fax: 49 6221 387 518. E-mail: akhtar@embl.de. Mailing address for Dirk Schübeler: Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland. Phone: 41 61 697 8269. Fax: 41 61 697 3976. E-mail: dirk@fmi.ch.

† Equally contributing authors.

<sup>∇</sup> Published ahead of print on 17 March 2008.

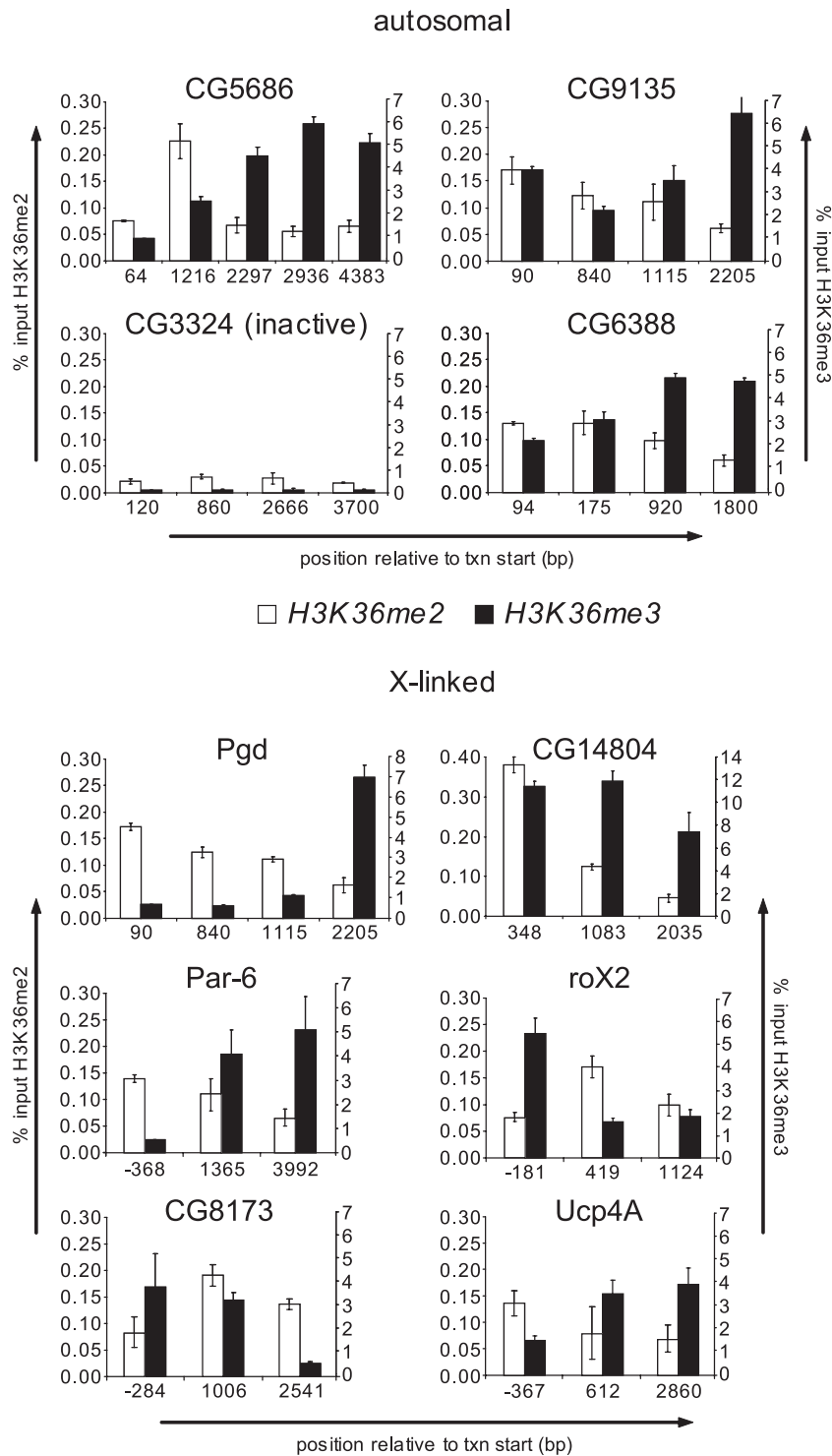


FIG. 1. High-resolution analysis of di- and trimethylation of H3K36 on autosomal and dosage-compensated genes. ChIP analysis of *Drosophila* SL2 cells using antibodies specific for H3K36me2 or H3K36me3 and quantification by real-time PCR. Shown are the average and standard deviation of ChIP enrichments from at least three independent experiments normalized to histone H3 occupancy. The *x* axis reflects the base pair position relative to the transcriptional start site. The *y* axis reflects enrichment (bound/input values are displayed as percent recovery of input DNA). H3K36me2, left scale; H3K36me3, right scale. Numbers in graphs are gene identification numbers according to Flybase.

ferase Hypb, similar to active autosomal genes (4). Despite comparable regulation, decreased H3K36me3 has an X-specific effect on the acetylation of H4 lysine 16 (H4K16ac), as it causes a reduction of that mark on dosage-compensated genes,

while on autosomal genes, levels are increased. Hypoacetylation on the male X chromosome as a consequence of Hypb loss of function coincides with reduced binding of the MSL1 and MOF proteins. Importantly, compromised MSL recruitment



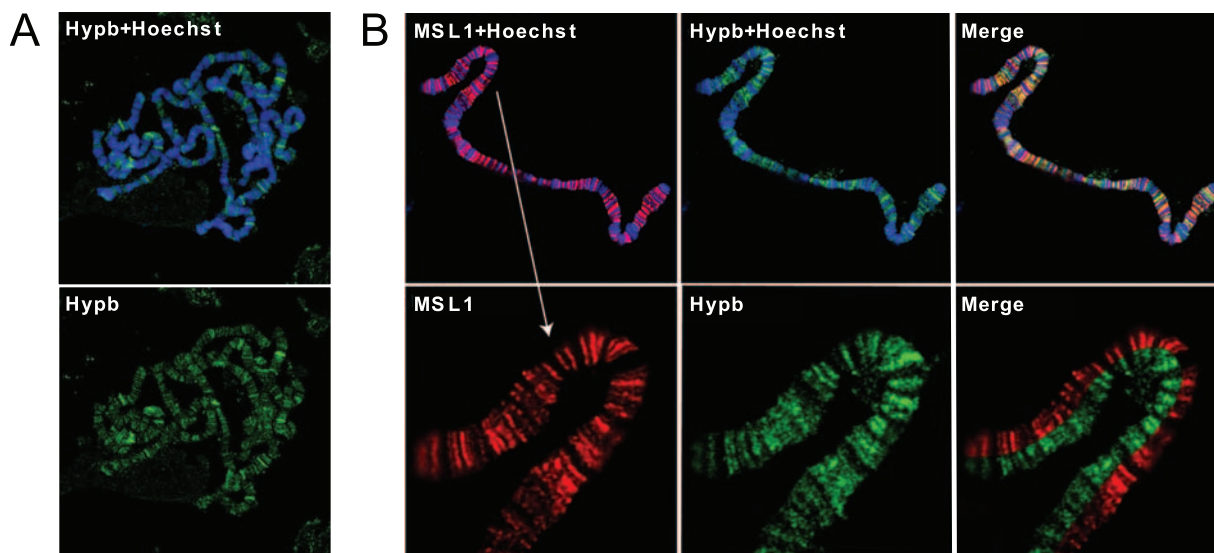


FIG. 2. Hybp binds autosomes and the male X chromosome. Polytene chromosomes of male third-instar larvae were stained with antibodies against Hybp (green) and MSL1 (red). DNA was visualized by Hoechst staining (blue). (A) Hybp localized preferentially to euchromatic interband regions, suggesting a general role in transcription. (B) Hybp partially localizes to sites of MSL1 enrichment along the X chromosome, consistent with a role in the trimethylation of H3K36 on dosage-compensated genes.

results in a failure to adequately upregulate the expression of a subset of X-linked genes. Thus, our data indicate that H3K36 trimethylation provides an important signal to attract MSL complex proteins to genes and further establish that the histone acetylation readout of H3K36 methylation in males is chromosome specific.

#### MATERIALS AND METHODS

**Tissue culture of SL2 cells.** *Drosophila* SL2 cells were kept in Schneider medium (Gibco) supplemented with 10% fetal calf serum.

**RNAi in cultured SL2 cells.** Double-stranded RNA (dsRNA) for RNAi knock-down of *Drosophila* Hybp (bp 3236 to 3944) was generated according to Ambion MEGAscript manual instructions.  $1 \times 10^6$  SL2 cells were plated in 2 ml medium and treated with 70  $\mu$ g dsRNA for 4 days. Treatment was repeated after cell splitting for a total of 7 days before harvesting cells for subsequent analysis.

**Western blot analysis and antibodies.** Western blottings were performed as previously described (4). Mouse monoclonal antibody against Hybp was used as previously described (4). Purified, bacterially expressed protein fragments were used to generate pMal-Hybp (amino acids [aa] 1 to 436), pMal-Hybp (aa 919 to 1135), and pMal-Hybp (aa 2040 to 2363), according to standard procedures. Hsp70 (mouse monoclonal; StressGen), H2A (Upstate 07-146), H3 (Abcam ab1791), H3K36me2 (Upstate 07-369), H3K36me3 (Abcam ab9050), H4K8ac (Upstate 07-328), H4K12ac (Upstate 07-595), H4K16ac (Upstate 07-329), MOF, and MSL1 (19) were used for the analysis.

**ChIP.** ChIPs of histone modifications, MOF and MSL1, were carried out as described previously (4).

**Immunostaining of polytene chromosomes.** Preparation of polytene chromosomes and immunostaining were performed as described previously (<http://www.igh.cnrs.fr/equip/cavalli/Lab%20Protocols/Immunostaining.pdf>). Hybp antibody and pre-serum were used in a 1:15 dilution; all other antibodies were used in a 1:250 dilution. Images were taken with a Leica Sp5 confocal microscope (Leica Microsystems, Mannheim) using an HCX PL APO 63.0 $\times$  1.40 oil objective.

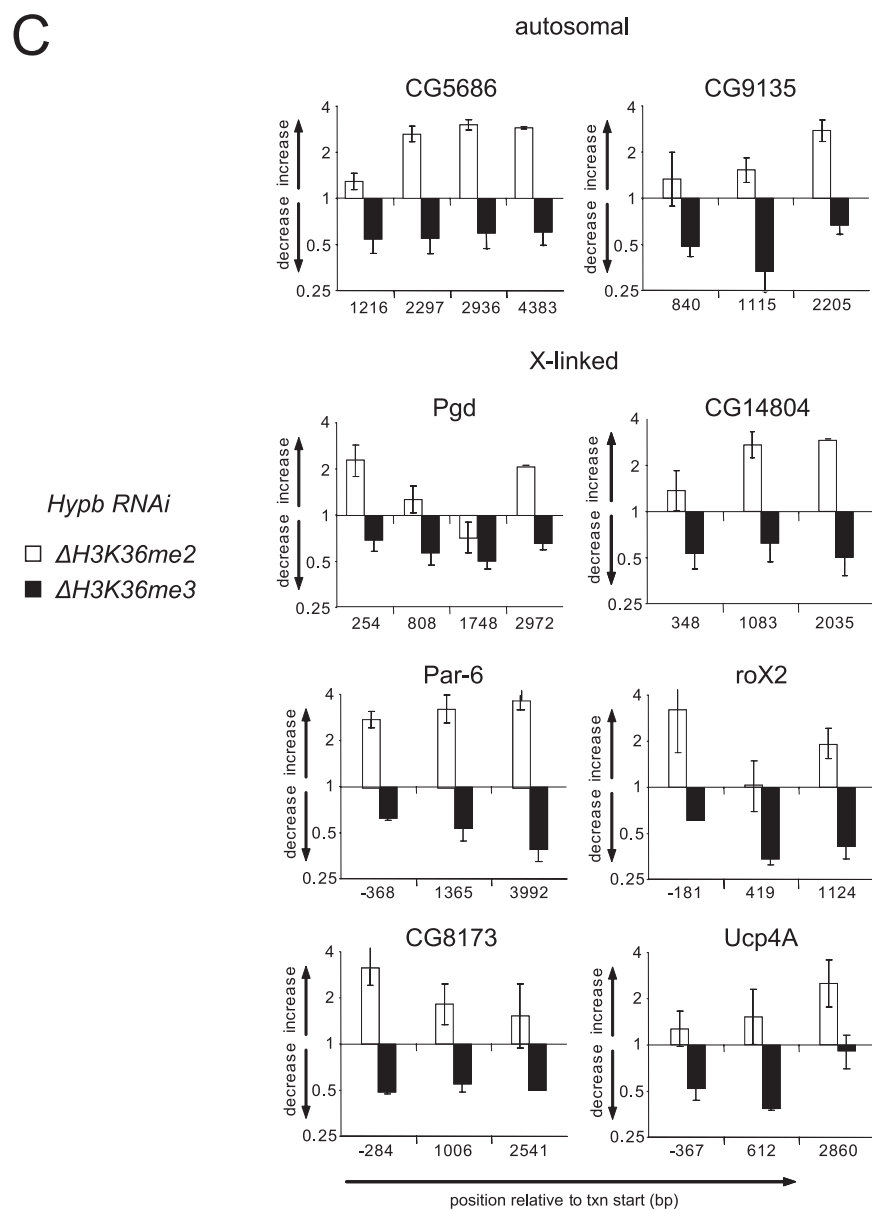
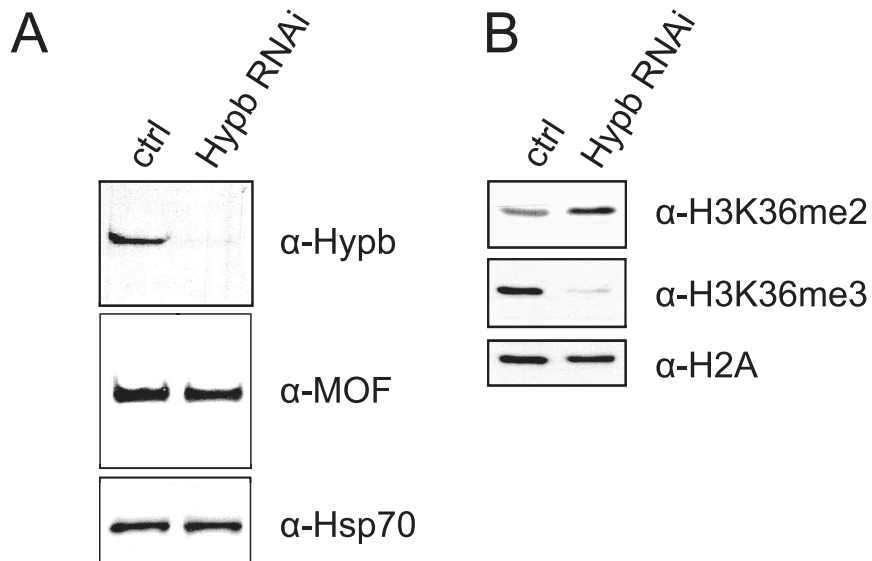
**Reverse transcription and real-time PCR.** Reverse transcription and quantitative real-time PCR analysis were performed as described previously (14). PCR conditions and autosomal primer sequences were as described previously (4, 28). Additional details for primer positions and sequences are available from the authors.

#### RESULTS

**Distributions of H3 lysine 36 methylation states are similar at dosage-compensated and autosomal genes.** To determine whether distribution of H3K36 methylation parallels the pattern of MSL binding, we performed ChIP with antisera specific for di- and trimethylation of this residue in male SL2 cells. ChIP enrichments were quantified by real-time PCR to determine K36 methylation states along selected X-linked and autosomal genes. On the X chromosome, both methylation states were enriched on a set of genes known to be subject to dosage compensation (14, 24, 26) yet displayed distinct patterns of localization. Dimethylation was preferentially localized proximal to promoters, whereas trimethylation peaked in the middle and in the 3' ends of dosage-compensated genes (Fig. 1), reminiscent of MSL binding. To determine whether K36 methylation states are distinct at MSL target genes, we compared X-linked profiles with di- and trimethylation on autosomes. Both methylation states were enriched along the body of transcriptionally active autosomal genes and assumed localizations that generally mirrored those of dosage-compensated genes (Fig. 1). This is in agreement with the distributions that we obtained recently for individual autosomal genes as well as for chromosome-wide profiles in female Kc cells (4).

These results indicate that the H3K36 methylation signature at dosage-compensated genes on the male X chromosome is similar to that on autosomes.

**Hybp mediates transcription-dependent trimethylation of H3 lysine 36 on autosomes and X chromosome.** In *Drosophila*, trimethylation of H3K36 relies on the activity of the SET-domain-containing protein CG1716 (4, 15, 24a, 29). Based on homology to the human histone methyltransferase HYPB (27), we refer to CG1716 as Hybp. In female cells, Hybp is enriched on actively transcribed genes and coincides with H3K36 tri-



methylation downstream of promoters (4). We determined the chromosomal distributions of this enzyme in *Drosophila* males by immunofluorescence staining of polytene chromosome squashes isolated from third-instar larvae. At low resolution, Hybp displayed binding to interbands and puffs with no apparent preference for any particular chromosome (Fig. 2A). When we performed costaining with an antibody against MSL1, we observed partial colocalization at many sites along the X chromosome (Fig. 2B). Despite an extensive overlap with MSL1 binding, we did not detect a characteristic pattern of Hybp localization on the compensated X chromosome, suggesting that Hybp is present at all sites of active transcription.

To define the contribution of Hybp to H3K36 trimethylation, we reduced transcript levels by RNAi in male SL2 cells. Hybp knockdown reduced protein levels, as indicated by Western blot analysis (Fig. 3A), and resulted in a specific decrease in H3K36 trimethylation, while dimethylation was slightly increased (Fig. 3B). These bulk methylation changes are similar to the ones previously observed following RNAi knockdown in female Kc cells (4).

Next, we examined by ChIP if the bulk effects on H3K36 methylation recapitulate changes at individual loci on autosomes and the X chromosome. At two autosomal genes, knockdown of Hybp reduced the presence of trimethylation and coincided with an increase of H3K36 dimethylation (Fig. 3C). This reflected the RNAi effects on bulk methylation and was similar to changes detected for individual genes in female Kc cells (4). Moreover, we show that the same chromatin changes also occur at six dosage-compensated genes following Hybp RNAi in male SL2 cells (Fig. 3C).

We conclude that Hybp mediates trimethylation of H3K36 on autosomes and the X chromosome in male *Drosophila* cells.

**H3K36 trimethylation is required for hyperacetylation of H4K16 on the dosage-compensated X chromosome.** Having established that H3K36me3 is equally regulated on all chromosomes, we asked if trimethylation mediates distinct downstream effects at autosomal and X-linked genes. In budding yeast, H3K36 methylation has been reported to regulate acetylation of histones H3 and H4 on transcribed open reading frames through recruitment of an HDAC-containing complex (5, 11, 13). Reduction of H3K36 trimethylation in female *Drosophila* cells also affected histone acetylation levels in transcribed regions, yet only for H4 lysine 16 acetylation (4). We therefore investigated if H4K16ac is subject to differential regulation in male cells. ChIP in male SL2 cells revealed that H4K16ac peaks at promoters of active autosomal genes and is less abundant along gene bodies (see the supplemental data at <http://www.fmi.ch/groups/schubeler.d/web/data.html>) (J. Kind, J. M. Vaquerizas, and A. Akhtar, unpublished data), similar to female Kc cells. In contrast, we detected highly elevated levels of H4K16ac along the gene body of dosage-compensated

genes, which is in agreement with previous reports (9, 24). On these genes, H4K16ac was especially abundant in the 3' end, reminiscent of the localization of H3K36me3 (see the supplemental data at <http://www.fmi.ch/groups/schubeler.d/web/data.html>).

Western blot analysis of Hybp knockdown showed that a reduction of trimethylation coincided with decreased levels of bulk acetylation at H4K16 (Fig. 4A). This was different from female cells where global acetylation increases in response to Hybp knockdown (4). Interestingly, when tested at specific loci by ChIP, we observed that levels of acetylation increased at autosomal genes yet at the same time decreased at dosage-compensated genes (Fig. 4B). We reasoned that this X-specific decrease is likely to account for the global reduction, since dosage-compensated genes show very high levels of H4K16 hyperacetylation compared to autosomal genes (see the supplemental data at <http://www.fmi.ch/groups/schubeler.d/web/data.html>) (24).

We conclude that while the presence of H3K36me3 reduces H4K16ac on autosomes, similar to female Kc cells, it has an additional male-specific function in enhancing H4K16ac on the dosage-compensated X chromosome. One possibility is that it contributes to MSL recruitment, which has previously been shown to be required for H4K16 hyperacetylation by MOF.

**H3K36 trimethylation enhances recruitment of MSL proteins at dosage-compensated genes.** The effect on H4K16ac suggests a reduction in MOF levels at target sites, indicating a function of H3K36me3 in MSL complex recruitment. To address this question, we reduced Hybp-dependent trimethylation and examined the levels of MSL recruitment in the same set of X-linked genes. We determined the binding pattern of two selected MSL complex members by ChIP using specific antibodies directed against MSL1 and MOF. Both proteins were bound along gene bodies, with the strongest enrichment found in the 3' ends, reminiscent of H3K36me3 (Fig. 5A) and in agreement with previous studies (2, 9). The addition of Hybp dsRNA and the subsequent reduction of H3K36me3 strongly diminished the presence of MOF at the four target genes assayed (Fig. 5A), which was consistent with the reduction in H4K16ac at these genes. Interestingly, while MSL1 abundance was also reduced at *Par-6*, *CG8173*, and *Ucp4A*, it remained mostly unaffected at sites of the *roX2* gene. The *roX2* gene presents 1 of 30 to 40 high-affinity sites, which contain sequence elements that are able to attract MSL1 to the X chromosome even in the absence of complete dosage compensation complexes (12) or transcription (14). The fact that MSL1 binding at the *roX2* locus is mostly insensitive to Hybp RNAi indicates that strong sequence affinity can mediate robust recruitment independent of H3K36me3. However, this MSL1 interaction appears to be insufficient to recruit a fully functional MSL complex, as reflected by reduced MOF

FIG. 3. RNAi knockdown of Hybp has similar effects at autosomal and X-linked loci. (A) Western blot analysis using antibody specific for Hybp reveals efficient reduction of Hybp in male *Drosophila* SL2 cells. Hsp70 and MOF remain unaffected by RNAi knockdown and serve as loading controls. (B) Reduction of Hybp results in a reduction of H3K36me3 and a coinciding increase of H3K36me2. H2A serves as a loading control. (C) Levels of H3K36 methylation states in RNAi and control cells were compared by ChIP followed by real-time PCR analysis. Shown is the ratio of H3K36me enrichments (change [*n*-fold], y axis) of RNAi over control cells relative to the position from the transcription start site (x axis). Effects on H3K36 methylation states at individual loci reflect bulk changes upon Hybp knockdown.

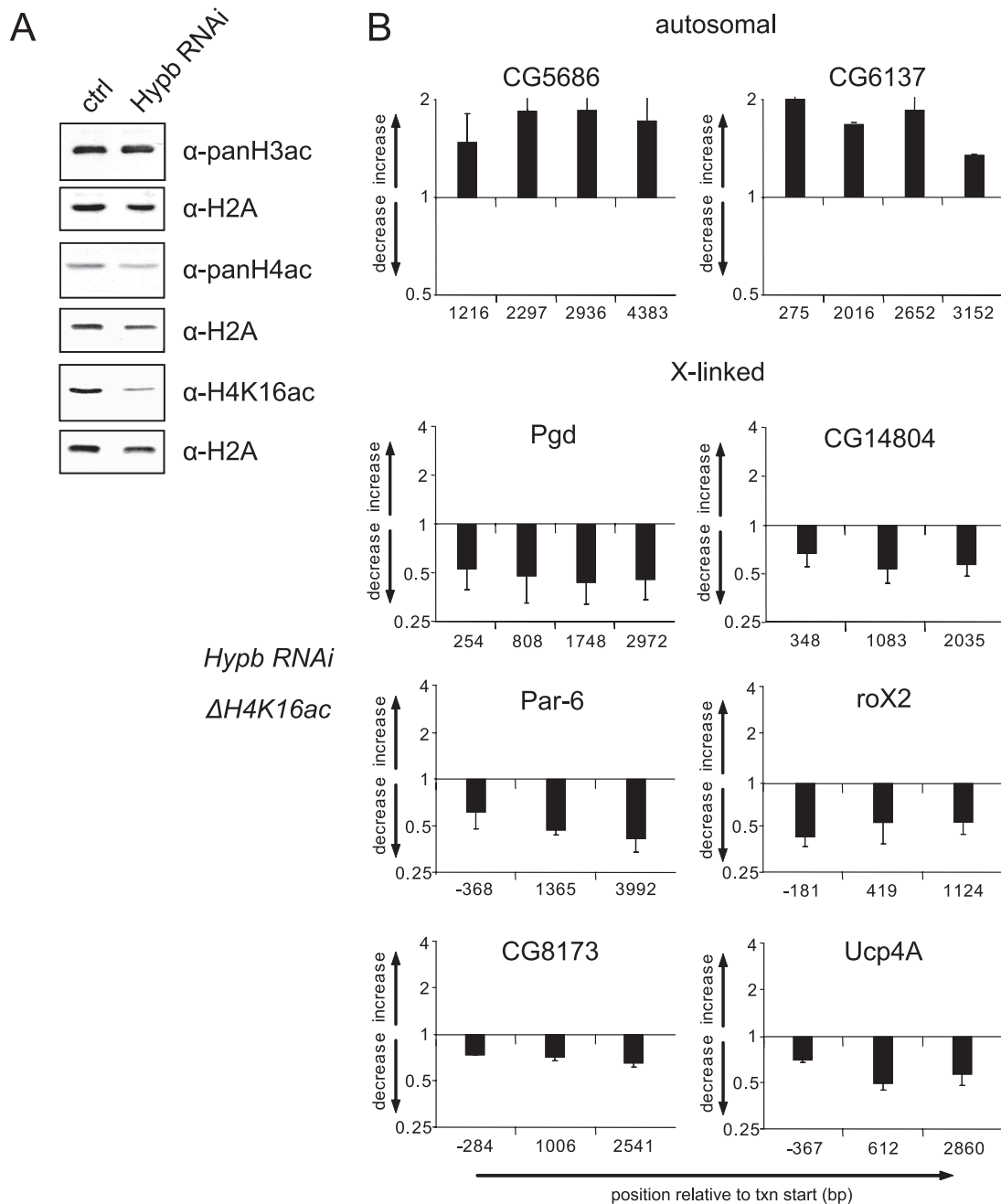


FIG. 4. H3K36 trimethylation is required for H4K16 hyperacetylation of the dosage-compensated X chromosome. (A) Hypb RNAi results in a reduction of bulk H4K16ac in male SL2 cells as indicated by Western blot analysis. (B) Comparison of changes in H4K16ac along autosomal and dosage-compensated X-linked genes upon RNAi by ChIP and real-time PCR. At autosomal genes, H4K16ac levels increase upon reduction of Hypb, whereas levels at dosage-compensated genes decrease.

binding and H4K16ac at the *roX2* gene under these conditions (Fig. 4B).

Hence, our results suggest that H3K36me3 is an important signal for stable association of a fully functional MSL complex with high-affinity sites and crucial for the binding of MSL proteins to low-affinity target genes. In the absence of trimethylation, MSL binding is diminished, resulting in lower levels of H4K16 hyperacetylation.

**K36 methylation is required for transcriptional regulation at MSL target genes.** Acetylation of H4K16 can relieve chromatin-mediated repression of transcription in vitro (1) and thus directly contribute to transcriptional upregulation of dosage-compensated genes. To address whether H3K36-dependent changes in acetylation affect transcription at target genes, we measured mRNA expression in untreated and Hypb knock-down cells by reverse transcription and quantitative real-time

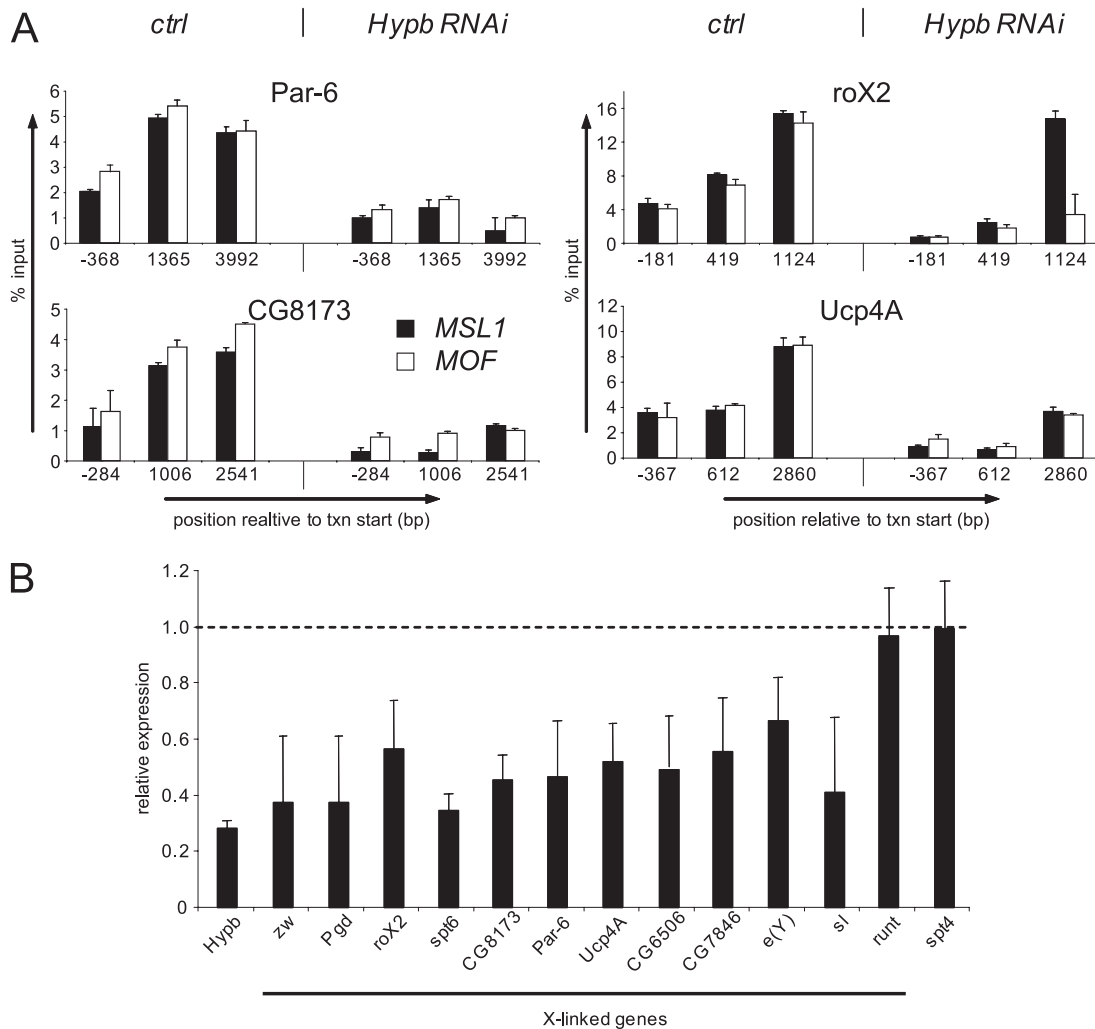


FIG. 5. H3K36 trimethylation enhances binding of the MSL complex and transcriptional upregulation at dosage-compensated genes. (A) MOF and MSL1 enrichments at dosage-compensated genes were compared with untreated and *Hypb* RNAi cells by ChIP and real-time PCR analysis. In control cells, both proteins displayed a similar pattern, localizing preferentially in the 3' end of genes. *Hypb* knockdown led to diminished levels of MOF and MSL1 at all positions along *Par-6*, *CG8173*, and *Ucp4A*. MOF was also depleted from high-affinity sites along the *roX2* gene, while the level of MSL1 remained largely unaffected. (B) Relative mRNA expression from dosage-compensated and noncompensated genes after *Hypb* knockdown. Displayed are average mRNA levels of five independent experiments, normalized to a mitochondrial RNA, comparing untreated and *Hypb* knockdown cells by quantitative real-time PCR (14). Upon *Hypb* RNAi, mRNA expression levels of 11 dosage-compensated genes are reduced approximately twofold. Expression levels of noncompensated X-linked *runt* as well as autosomal *spt4* remain unchanged.

PCR. Indeed, expression of all tested dosage-compensated genes was significantly reduced upon decline of H3K36me3 (Fig. 5B). This effect was similar to the approximately twofold decrease of target gene mRNA levels after MSL2 knockdown (26), indicating that trimethylation is important for adequate transcriptional upregulation of X-linked genes. However, expression was not similarly decreased at all genes tested, since an X-linked gene, which is not subject to dosage compensation, and an autosomal gene remained unaffected. These results emphasize the critical role of H3K36me3 as a chromatin signature to allow recruitment of MSL proteins to sites of transcriptional compensation.

## DISCUSSION

In this study, we report that trimethylation of histone H3 lysine 36 is required for high levels of H4K16ac at dosage-

compensated genes on the male X chromosome. This function does not reflect an X-specific methylation signature, since both H3K36 methylation states have similar localization patterns at autosomal genes: dimethylation peaks promoter-proximal, and trimethylation shows a 3' bias (Fig. 1). Furthermore, the regulation of H3K36me3 depends on the activity of *Hypb*, which is equally targeted to autosomal and X-linked loci, indicating a common mode of regulation (Fig. 2 and 3).

Nevertheless, downregulation of H3K36me3 in *Drosophila* SL2 cells resulted in reduced levels of H4K16 hyperacetylation at X-linked genes but simultaneously increased levels at autosomal genes in the same cells (Fig. 4). This differential effect on acetylation suggests a context-dependent readout of lysine 36 methylation. In *Saccharomyces cerevisiae*, H3K36me signals binding of the chromodomain-containing protein Eaf3, which in turn recruits an Rpd3 complex to deacetylate the 3' end of



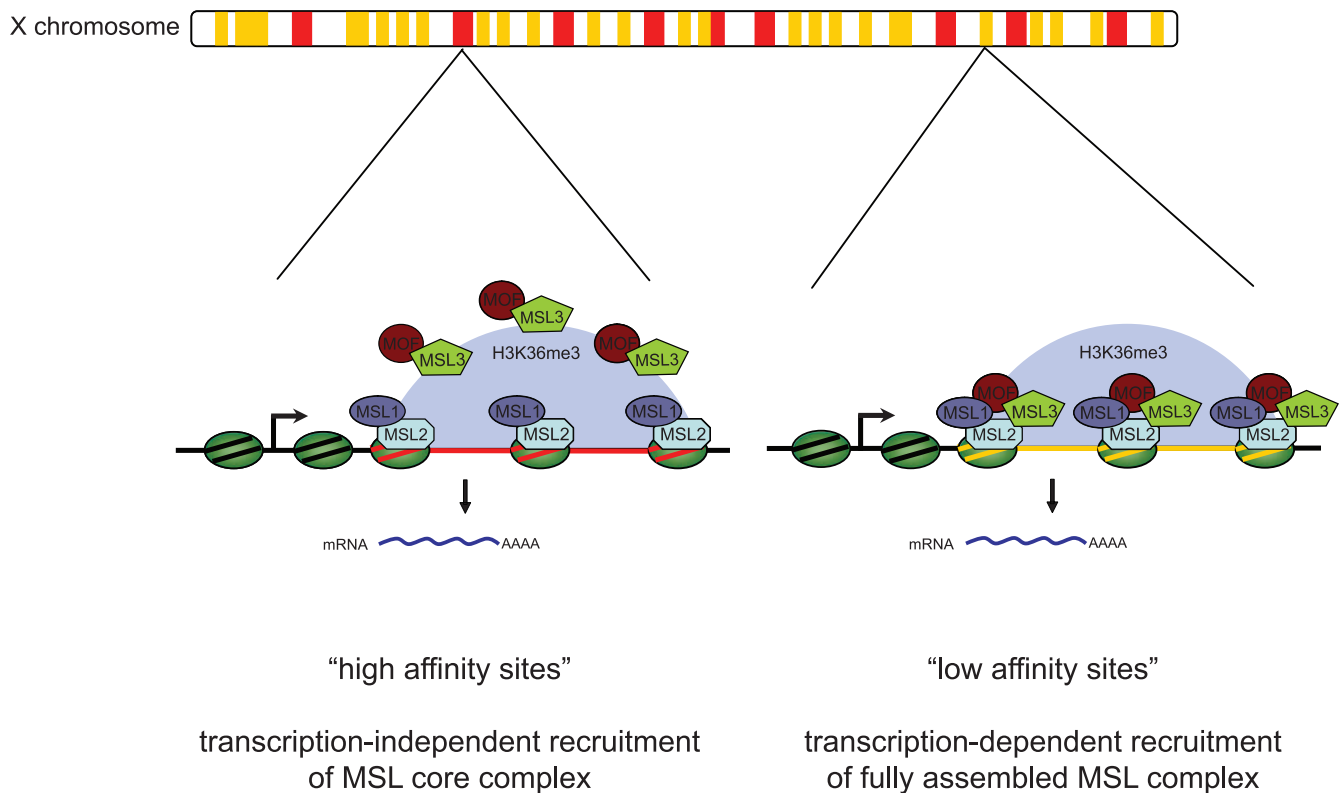


FIG. 6. Model for MSL complex targeting to sites of dosage compensation in *Drosophila*. Genes along the X chromosome have various affinities for MSL complex binding. High-affinity sites (red) can attract partially assembled MSL complexes independent of transcription-coupled chromatin modifications. However, H3K36me3 is still necessary to facilitate robust interaction with MOF and MSL3. In comparison, the majority of target genes contain promoter-distal sequence elements, which have relatively weak affinity for MSL recruitment (low-affinity sites [yellow]). At these genes, transcription-dependent H3K36me3 (blue arch represents concentration of H3K36me3) enhances recognition and stable binding of fully assembled MSL complexes. Thus, robust recruitment of the MSL proteins to the dosage-compensated X chromosome relies on combined contributions of degenerate sequence elements and transcription-coupled histone modifications.

transcribed genes (5, 11, 13). We provide evidence that the X-specific reduction of histone acetylation in Hybp-depleted *Drosophila* SL2 cells reflects compromised recruitment of MSL1 and MOF at dosage-compensated genes (Fig. 5A). This is in full agreement with reduced binding of MSL3 upon Hybp knockdown, which was recently reported by Larschan and colleagues (15). MSL3 is one of the *Drosophila* homologues of yeast Eaf3 (7) and localizes together with MOF and MSL1 to the 3' end of dosage-compensated genes (2, 9). Thus, in analogy to yeast, MSL3 is likely to associate with H3K36me3 at the 3' end of X-linked genes, leading to robust complex binding and enhanced H4K16 hyperacetylation. This is supported by evidence showing that MSL3 preferentially interacts with Set2-methylated nucleosomes in vitro (15). Moreover, our observation of Hybp localizing to active sites on polytene chromosomes provides further evidence for a direct role of H3K36me3 in MSL recruitment (Fig. 2B). However, not all sites enriched for Hybp were also bound by MSL1, suggesting that H3K36me3 is necessary but not sufficient for MSL complex recruitment.

Whereas proper binding of the MSL complex to *Par-6*, *CG8173*, and *Ucp4A* relies on the presence of H3K36me3, Hybp knockdown did not significantly decrease MSL1 recruitment at the *roX2* gene (Fig. 5A) (15). This is similar to the

binding of MSL1 and MSL2 to high-affinity sites in *msl3* or *mof* mutant flies (10, 12, 17), suggesting that strong sequence affinity can target partial MSL complexes independent of H3K36me3. Importantly, despite its presence at the *roX2* locus in Hybp knockdown cells, MSL1 was insufficient for adequate MOF recruitment and transcriptional upregulation (Fig. 5A and B). Thus, our data indicate that H3K36me3 is necessary at high-affinity sites to facilitate robust MOF interaction and the subsequent hyperacetylation needed to double transcription (Fig. 6).

Interestingly, *roX2* transcription was unaffected by Hybp RNAi when expressed from a plasmid model system (29). Since the consequence of reduced H3K36me3 on H4K16ac on the *roX2* plasmid was not determined in this study, it is possible that a less pronounced reduction in acetylation might account for this effect.

In contrast to the *roX2* gene, H3K36me3 was required for MSL1 binding to lower-affinity genes. At these genes, transcription-dependent methylation might facilitate DNA accessibility in the 3' end by enhancing the recruitment of MOF and the hyperacetylation of H4K16 (Fig. 6).

At autosomal genes, reduced trimethylation caused the opposite effect on H4 lysine 16 acetylation. Thus, one modification may signal two different outcomes in the same cell in a

chromosome-specific fashion. It is conceivable that such differential readouts involve interaction with either distinct methyl-binding proteins or alternative subunit compositions.

The presence of antagonistic activities in the same nucleus, which are targeted to the same modification, requires spatial restriction of individual protein complexes to avoid deregulation by improper acetylation or deacetylation. Thus, the preferential interaction of MSL proteins with H3K36me3 on the X chromosome might be favored by locally accumulating MSL proteins at high-affinity sites. MSL interactions with nuclear pore proteins (19) suggest a possible role of nuclear organization in X chromosome dosage compensation, which may further contribute to a preferential binding of MSL proteins to H3K36me3. Conversely, while this confines histone acetyltransferase activity to dosage-compensated genes on the X chromosome, it might also ensure that the same activity is not mistargeted to autosomal genes.

#### ACKNOWLEDGMENTS

We thank the FMI antibody facility for the generation of antisera (Susanne Schenk) and members of the Akhtar and Schübeler labs for advice during the course of the project and comments on the manuscript.

A.A. thanks Leica Microsystems for continuous support of the Advanced Light Microscopy Facility. We further acknowledge support by the Novartis Research Foundation (D.S.), DFG SPP1129 "Epigenetics" (A.A.), and the EU 6th framework program (NOE "The Epigenome" LSHG-CT-2004-503433 to D.S. and A.A. and LSHG-CT-2006-037415 to D.S.).

#### REFERENCES

1. Akhtar, A., and P. B. Becker. 2000. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in *Drosophila*. *Mol. Cell* 5:367–375.
2. Alekseyenko, A. A., E. Larschan, W. R. Lai, P. J. Park, and M. I. Kuroda. 2006. High-resolution ChIP-chip analysis reveals that the *Drosophila* MSL complex selectively identifies active genes on the male X chromosome. *Genes Dev.* 20:848–857.
3. Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837.
4. Bell, O., C. Wirbelauer, M. Hild, A. N. Scharf, M. Schwaiger, D. M. MacAlpine, F. Zilbermann, F. van Leeuwen, S. P. Bell, A. Imhof, D. Garza, A. H. Peters, and D. Schubeler. 2007. Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in *Drosophila*. *EMBO J.* 26:4974–4984.
5. Carrozza, M. J., B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W. J. Shia, S. Anderson, J. Yates, M. P. Washburn, and J. L. Workman. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123:581–592.
6. Dahlsveen, I. K., G. D. Gilfillan, V. I. Shelest, R. Lamm, and P. B. Becker. 2006. Targeting determinants of dosage compensation in *Drosophila*. *PLoS Genet.* 2:e5.
7. Eisen, A., R. T. Utley, A. Nourani, S. Allard, P. Schmidt, W. S. Lane, J. C. Lucchesi, and J. Cote. 2001. The yeast NuA4 and *Drosophila* MSL complexes contain homologous subunits important for transcription regulation. *J. Biol. Chem.* 276:3484–3491.
8. Fagiegaltier, D., and B. S. Baker. 2004. X chromosome sites autonomously recruit the dosage compensation complex in *Drosophila* males. *PLoS Biol.* 2:e341.
9. Gilfillan, G. D., T. Straub, E. de Wit, F. Greil, R. Lamm, B. van Steensel, and P. B. Becker. 2006. Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex. *Genes Dev.* 20:858–870.
10. Gu, W., P. Szauter, and J. C. Lucchesi. 1998. Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of *Drosophila melanogaster*. *Dev. Genet.* 22:56–64.
11. Joshi, A. A., and K. Struhl. 2005. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol. Cell* 20:971–978.
12. Kelley, R. L., V. H. Meller, P. R. Gordadze, G. Roman, R. L. Davis, and M. I. Kuroda. 1999. Epigenetic spreading of the *Drosophila* dosage compensation complex from roX RNA genes into flanking chromatin. *Cell* 98:513–522.
13. Keogh, M. C., S. K. Kurdastani, S. A. Morris, S. H. Ahn, V. Podolny, S. R. Collins, M. Schuldiner, K. Chin, T. Punna, N. J. Thompson, C. Boone, A. Emili, J. S. Weissman, T. R. Hughes, B. D. Strahl, M. Grunstein, J. F. Greenblatt, S. Buratowski, and N. J. Krogan. 2005. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* 123:593–605.
14. Kind, J., and A. Akhtar. 2007. Cotranscriptional recruitment of the dosage compensation complex to X-linked target genes. *Genes Dev.* 21:2030–2040.
15. Larschan, E., A. A. Alekseyenko, A. A. Gortchakov, S. Peng, B. Li, P. Yang, J. L. Workman, P. J. Park, and M. I. Kuroda. 2007. MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol. Cell* 28:121–133.
16. Lucchesi, J. C., W. G. Kelly, and B. Panning. 2005. Chromatin remodeling in dosage compensation. *Annu. Rev. Genet.* 39:615–651.
17. Lyman, L. M., K. Copps, L. Rastelli, R. L. Kelley, and M. I. Kuroda. 1997. *Drosophila* male-specific lethal-2 protein: structure/function analysis and dependence on MSL-1 for chromosome association. *Genetics* 147:1743–1753.
18. Mendjan, S., and A. Akhtar. 2007. The right dose for every sex. *Chromosoma* 116:95–106.
19. Mendjan, S., M. Taipale, J. Kind, H. Holz, P. Gebhardt, M. Schelder, M. Vermeulen, A. Buscaino, K. Duncan, J. Mueller, M. Wilm, H. G. Stunnenberg, H. Saumweber, and A. Akhtar. 2006. Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Mol. Cell* 21:811–823.
20. Mikkelsen, T. S., M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E. S. Lander, and B. E. Bernstein. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448:553–560.
21. Reference deleted.
22. Pokholok, D. K., C. T. Harbison, S. Levine, M. Cole, N. M. Hannett, T. I. Lee, G. W. Bell, K. Walker, P. A. Rolfe, E. Herbolzheimer, J. Zeitlinger, F. Lewitter, D. K. Gifford, and R. A. Young. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122:517–527.
23. Rao, B., Y. Shibata, B. D. Strahl, and J. D. Lieb. 2005. Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Mol. Cell. Biol.* 25:9447–9459.
24. Smith, E. R., C. D. Allis, and J. C. Lucchesi. 2001. Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in *Drosophila* males. *J. Biol. Chem.* 276:31483–31486.
- 24a. Stabell, M., J. Larsson, R. B. Aalen, and A. Lambertsson. 2007. *Drosophila* dSet2 functions in H3-K36 methylation and is required for development. *Biochem. Biophys. Res. Commun.* 359:784–789.
25. Straub, T., and P. B. Becker. 2007. Dosage compensation: the beginning and end of generalization. *Nat. Rev. Genet.* 8:47–57.
26. Straub, T., G. D. Gilfillan, V. K. Maier, and P. B. Becker. 2005. The *Drosophila* MSL complex activates the transcription of target genes. *Genes Dev.* 19:2284–2288.
27. Sun, X. J., J. Wei, X. Y. Wu, M. Hu, L. Wang, H. H. Wang, Q. H. Zhang, S. J. Chen, Q. H. Huang, and Z. Chen. 2005. Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *J. Biol. Chem.* 280:35261–35271.
28. Wirbelauer, C., O. Bell, and D. Schubeler. 2005. Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. *Genes Dev.* 19:1761–1766.
29. Yokoyama, R., A. Pannuti, H. Ling, E. R. Smith, and J. C. Lucchesi. 2007. A plasmid model system shows that *Drosophila* dosage compensation depends on the global acetylation of histone H4 at lysine 16 and is not affected by depletion of common transcription elongation chromatin marks. *Mol. Cell. Biol.* 27:7865–7870.





## **Appendix B**

**“The nonspecific lethal complex is a transcriptional regulator in *Drosophila*.”**

Raja SJ, Charapitsa I, Conrad T, Vaquerizas JM, Gebhardt P, Holz H, Kadlec J, Fraterman S,  
Luscombe NM, Akhtar A.

Mol Cell. 2010 Jun 25;38(6):827-41.



# The Nonspecific Lethal Complex Is a Transcriptional Regulator in *Drosophila*

Sunil Jayaramaiah Raja,<sup>1,2,5</sup> Iryna Charapitsa,<sup>1,2,5</sup> Thomas Conrad,<sup>1,2</sup> Juan M. Vaquerizas,<sup>3</sup> Philipp Gebhardt,<sup>1</sup> Herbert Holz,<sup>1,2</sup> Jan Kadlec,<sup>4</sup> Sven Fraterman,<sup>1</sup> Nicholas M. Luscombe,<sup>1,3</sup> and Asifa Akhtar<sup>1,2,\*</sup>

<sup>1</sup>Genome Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

<sup>2</sup>Max Planck Institute of Immunobiology, Stübeweg 51, D-79108 Freiburg, Germany

<sup>3</sup>European Molecular Biology Laboratory European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK

<sup>4</sup>European Molecular Biology Laboratory Grenoble, 6, rue Jules Horowitz, BP 181, 38042 Grenoble Cedex 9, France

<sup>5</sup>These authors contributed equally to this work

\*Correspondence: [akhtar@immunbio.mpg.de](mailto:akhtar@immunbio.mpg.de)

DOI 10.1016/j.molcel.2010.05.021

## SUMMARY

Here, we report the biochemical characterization of the nonspecific lethal (NSL) complex (NSL1, NSL2, NSL3, MCERS2, MBD-R2, and WDS) that associates with the histone acetyltransferase MOF in both *Drosophila* and mammals. Chromatin immunoprecipitation-Seq analysis revealed association of NSL1 and MCERS2 with the promoter regions of more than 4000 target genes, 70% of these being actively transcribed. This binding is functional, as depletion of MCERS2, MBD-R2, and NSL3 severely affects gene expression genome wide. The NSL complex members bind to their target promoters independently of MOF. However, depletion of MCERS2 affects MOF recruitment to promoters. NSL complex stability is interdependent and relies mainly on the presence of NSL1 and MCERS2. Tethering of NSL3 to a heterologous promoter leads to robust transcription activation and is sensitive to the levels of NSL1, MCERS2, and MOF. Taken together, we conclude that the NSL complex acts as a major transcriptional regulator in *Drosophila*.

## INTRODUCTION

Eukaryotic genes that encode messenger RNAs are subject to primary regulation at the level of transcription. A series of distinct phases occur at the onset of transcription of these RNAs, beginning with the binding of activators upstream of the core promoter, followed by the recruitment of adaptor complexes such as SAGA or mediator. In turn those adaptor complexes facilitate the binding of general transcription factors (GTFs) and RNA polymerase II and initiate transcription (Thomas and Chiang, 2006). To better understand the mechanism of transcription initiation, characterization of yet-unidentified promoter-bound proteins is essential.

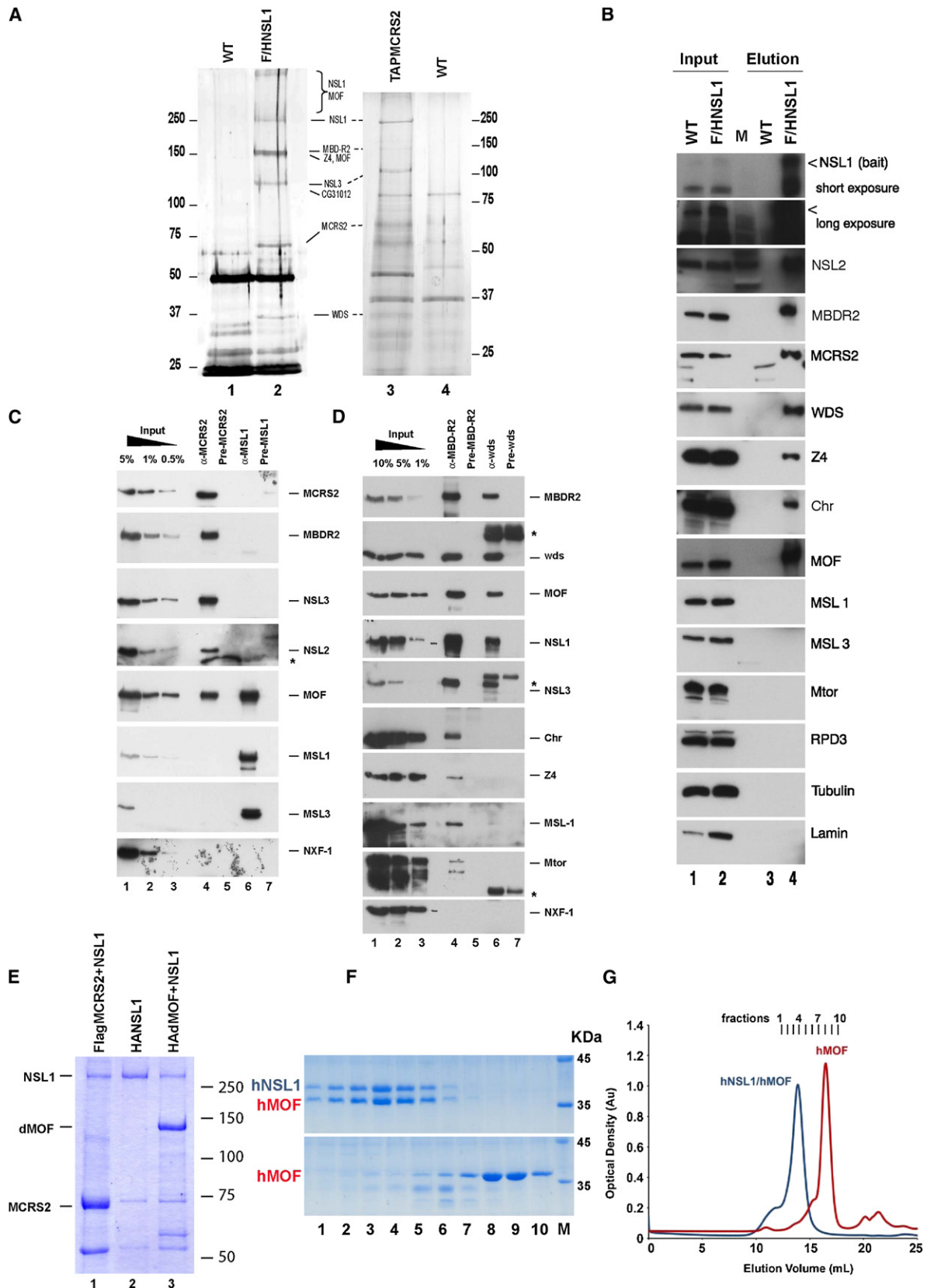
Transcription initiation in eukaryotes involves dynamic changes in chromatin structure that permit assembly of the

transcription machinery at a gene promoter (Lemon and Tjian, 2000; Orphanides and Reinberg, 2000). The fundamental structural unit of chromatin is the nucleosome, which contains 146 bp of DNA wrapped around a histone octamer composed of two copies each of histones H2A, H2B, H3, and H4. Histones in a nucleosomal context are subject to a variety of posttranslational modifications, such as acetylation, methylation, phosphorylation, ribosylation, ubiquitinylation, etc. A number of well-conserved enzymes carry out these modifications (for review, see Kouzarides, 2007).

Males absent on first (MOF) is a histone H4 lysine 16 specific acetyltransferase in both *Drosophila* and mammals (Hilfiker et al., 1997; Mendjan et al., 2006; Smith et al., 2000; Taipale et al., 2005). In *Drosophila*, MOF is well known for its role in dosage compensation of the male X chromosome in the context of the male specific lethal (MSL) complex (Akhtar and Becker, 2000; Hilfiker et al., 1997; Kind et al., 2008; Smith et al., 2000, 2001). Genome-wide high-resolution binding profiles of MOF along with MSL1 and MSL3 have shown that MSLs are enriched primarily toward the 3' end of X-linked genes. In contrast, MOF also binds to promoter proximal regions of the same genes as well as to a large number of autosomal promoters, independently of the other MSL complex members (Alekseyenko et al., 2006, 2008; Gilfillan et al., 2006; Kind et al., 2008; Legube et al., 2006). However, it remained unclear whether MOF binds to promoters alone or whether it is associated with additional proteins.

Our previous studies have shown that MOF associates not only with the MSL complex members, but also with a number of uncharacterized proteins such as CG4699 (NSL1), CG18041 (NSL2), CG8233 (NSL3), CG1135 (MCERS2), and CG10042 (MBD-R2). These proteins were named as nonspecific lethal (NSL) proteins since disruption of the respective genes by *P*-element insertions in *Drosophila* is early larval lethal in both sexes (Mendjan et al., 2006). However, it remained unknown whether there was any functional link between these proteins and MOF.

To gain further insight, in the present study, we performed the purification and functional characterization of TAP-tagged MCERS2 and TAP-HA-FLAG-tagged NSL1. We find NSL1, NSL2, NSL3, MCERS2, MBD-R2, WDS, and MOF consistently purified as a complex, which we name the NSL complex.



Interestingly, members of the NSL complex bind to MOF target promoters on the X chromosome and autosomes in both sexes. Chromatin immunoprecipitation (ChIP)-Seq profiling for NSL1 and MCRS2 reveals that these proteins together bind to the promoters of more than 4000 target genes. The NSL complex members bind to their target promoters independently of MOF. However, MOF targeting to promoters is NSL complex dependent, showing a hierarchy of recruitment. Furthermore, upon depletion of NSL1, NSL3, or MCRS2, the stability of the NSL complex is compromised. MCRS2, MBD-R2, and NSL3 are important regulators of gene expression, as their depletion severely affects the mRNA levels of most of the NSL target genes in a genome-wide manner. Our data suggest that reduction in transcript levels is the result of impaired transcription initiation, since it correlates with reduced levels of RNA polymerase II at gene promoters. In addition, we show that tethering NSL3 to a heterologous promoter activates transcription, and that the NSL complex members as well as MOF modulate this activity. These results suggest a cooperative interaction between the NSL complex members and MOF. Taken together, we identify the NSL complex as an evolutionarily conserved complex, which acts as a major transcriptional regulator in *Drosophila*.

## RESULTS

### Biochemical Purification of the NSL Complex

Our previous copurification of MOF interacting proteins identified a set of proteins of unknown function (Mendjan et al., 2006). To gain further insight into the nature of these interactions, we generated stable Schneider (SL-2) cell lines expressing two of the uncharacterized proteins—TAP-tagged MCRS2 and TAP-HA-FLAG-tagged NSL1. Nuclear extracts were prepared from cell lines that express tagged proteins as well as from wild-type cells for mock purification. The quality of the affinity-purified material was analyzed by SDS-PAGE followed by silver staining (Figure 1A). Matrix-assisted laser desorption/ionization time of flight, nanoelectrospray, liquid chromatography-tandem mass spectrometry, as well as western blot analysis identified the following proteins consistently purifying with either NSL1 or MCRS2: NSL1, NSL2, NSL3, MCRS2, MBD-R2, WDS, and MOF in addition to a number of other proteins (Figure 1A; Table S1, available online).

The *ns1* gene is located on chromosome 3R and encodes a protein of 1550 aa which contains a MOF interacting PEHE domain at its C terminus. The *ns2* gene is located on chromosome 3R and encodes a protein of 484 aa that contains two C/H-rich domains. The *ns3* gene is located on chromosome 2R and encodes three splice variant proteins of 1001 aa, 1066 aa, and 934 aa, all containing a  $\alpha/\beta$  hydrolase domain. The *mcrs2* gene is located on chromosome 3L and encodes a protein of 578 aa containing a ForkHead-Associated domain (FHA) at its C terminus. The *mbd-r2* gene is located on chromosome 3R and encodes two splice variants: one variant of 1169 aa containing a Tudor, MBD, ZnF and a PHD finger domain and a second variant of 1081 aa without the Tudor domain (Hendrich and Tweedie, 2003).

None of these proteins were found in the control mock purification from the wild-type SL-2 cell line, thus validating the purification assay. The interactions were further confirmed by western blot analysis of the eluted fractions (Figure 1B). Interestingly, apart from MOF, none of the other MSL complex proteins copurified with MCRS2 or NSL1. These results, as well as coimmunoprecipitation experiments with WDS, MBD-R2, MCRS2, and MSL1-specific antibodies revealed that the interaction between these proteins and MOF was specific and distinct from the previously characterized MSL complex (Figures 1C and 1D). Among the proteins tested in coimmunoprecipitation experiments only MBD-R2 revealed a substoichiometric interaction with MSL1. We termed this copurified complex containing NSL1, NSL2, NSL3, MCRS2, MBD-R2, WDS, and MOF as the NSL complex. NSL complex elutions were also tested for histone acetyltransferase (HAT) activity. Indeed, we detect an enrichment of histone H4 directed HAT activity in these fractions (Figures S1A and S1B).

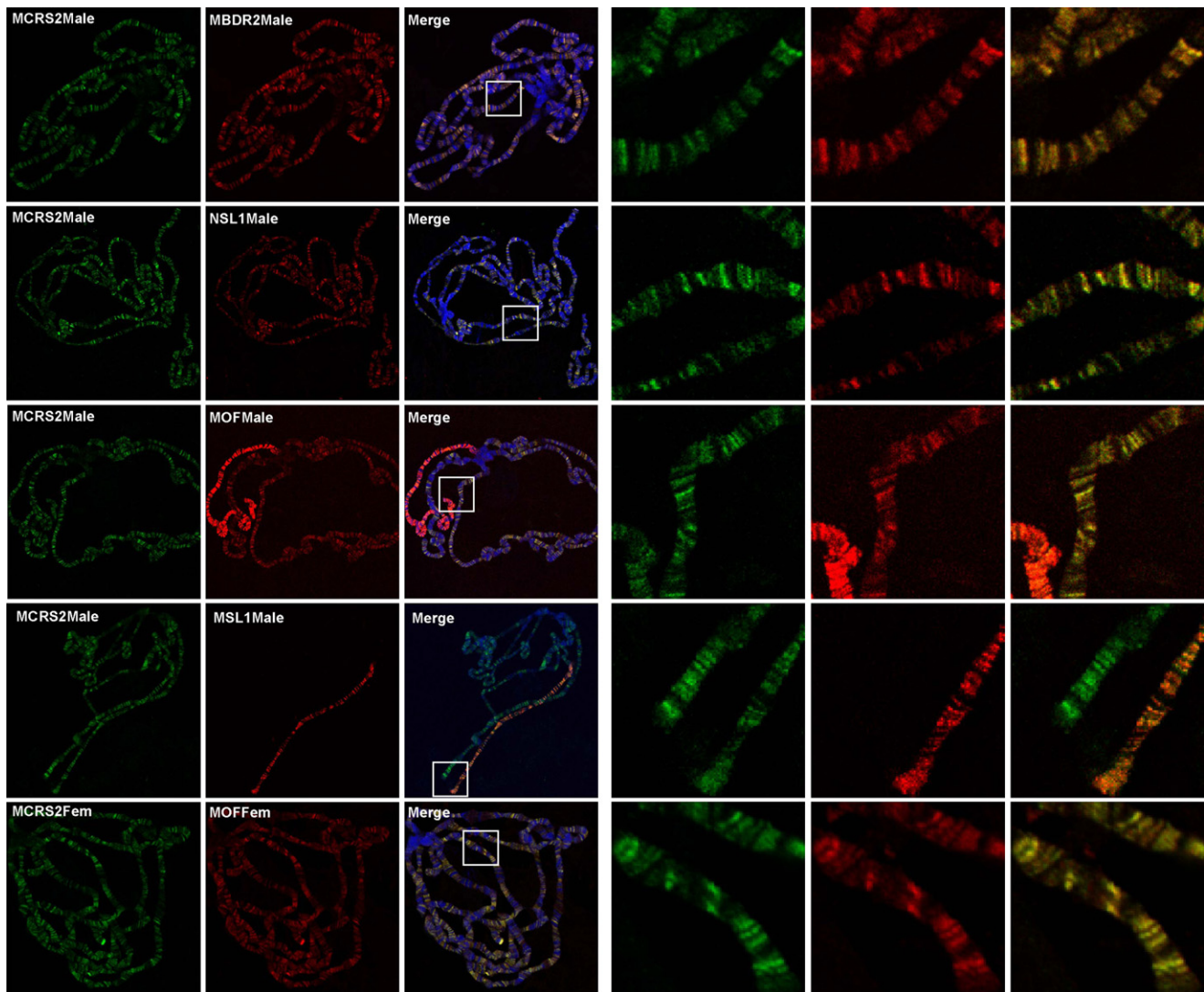
### NSL1 Directly Interacts with MCRS2 and MOF

In order to dissect the interactions between the NSL proteins, expression constructs with tagged (FLAG or HA) as well as untagged NSL1, MCRS2, and MOF were expressed in the baculovirus expression system (Figure 1E; data not shown). Copurification of these proteins revealed a stable interaction of NSL1 with MCRS2 and MOF (Figure 1E, lanes 1–3). These results showed that the interaction between NSL1-MCRS2 and NSL1-MOF is direct and can occur in the absence of other NSL

### Figure 1. Purification of the NSL1 and MCRS2 Complexes

- (A) Silver staining of copurified proteins from Schneider SL-2 cells stably expressing TAP-FLAG-HA-NSL1 (lanes 1 and 2) and TAP-MCRS2 (lanes 3 and 4). WT indicates corresponding mock purifications from wild-type SL-2 cells. From each cell line 1.5 ml of nuclear extract with a protein concentration of 6  $\mu\text{g}/\mu\text{l}$  was used, and 50% of the purified eluted material was loaded on a gel; the rest of the material was used for subsequent mass spectrometry analysis. The bracket indicates an additional area on the gel where NSL1 and MOF peptides were identified upon band excision by mass spectrometry analysis.
- (B) Western blot analysis of the TAP-FLAG-HA-NSL1 (TFH-NSL1) purification for NSL proteins. NSL1, NSL2, MCRS2, MBD-R2, WDS, Z4, Chromator, and MOF were detected in the final eluate, but not MSL1 and MSL3. Tubulin and lamin served as negative controls. The open arrow indicates the NSL1 bait. Long exposure is shown for detection of NSL1 in input lanes.
- (C) Immunoprecipitation from *Drosophila* embryo nuclear extract with  $\alpha$ -MCRS2 or  $\alpha$ -MSL1 antisera. The blot was probed with various antibodies as indicated. Asterisks represent the IgG band.
- (D) Same as in (C), except immunoprecipitations were performed with  $\alpha$ -MBDR2 or  $\alpha$ -WDS antibodies.
- (E) Reconstitution of NSL interactions using baculovirus-expressed proteins. Interaction of NSL1 with MCRS2 (lane 1) and *Drosophila* MOF (dMOF, lane 3) upon incubation of protein extracts. Purification of HA-tagged NSL1 alone (lane 2). After purification via the corresponding tag, proteins were run on a SDS-PAGE and the gel stained with Coomassie blue.
- (F) Fractions 1 to 10 of the hNSL1 (aa 883–1105)/hMOF (aa 174–458) (top) and hMOF (aa 174–458) (bottom) Superdex 200 gel filtration profiles. Fractions were run on a SDS-PAGE and the gel stained with Coomassie blue.
- (G) Superdex 200 gel filtration elution profiles of hNSL1 (aa 883–1105)/hMOF (aa 174–458) in blue and of hMOF (174–458) in red.





**Figure 2. MCRS2, MBD-R2, NSL1, and MOF Colocalize Globally on Wild-Type Male and Female *Drosophila* Third Instar Larval Polytene Chromosomes**

Coimmunofluorescence staining of MCRS2 (green) with MBD-R2, NSL1, MOF, and MSL1 (red); DNA counterstained by Hoechst. As a control of binding specificity, coimmunostaining of MCRS2 and MSL1 is shown. White boxes indicate the zoomed area in the right panels showing the extent of colocalization.

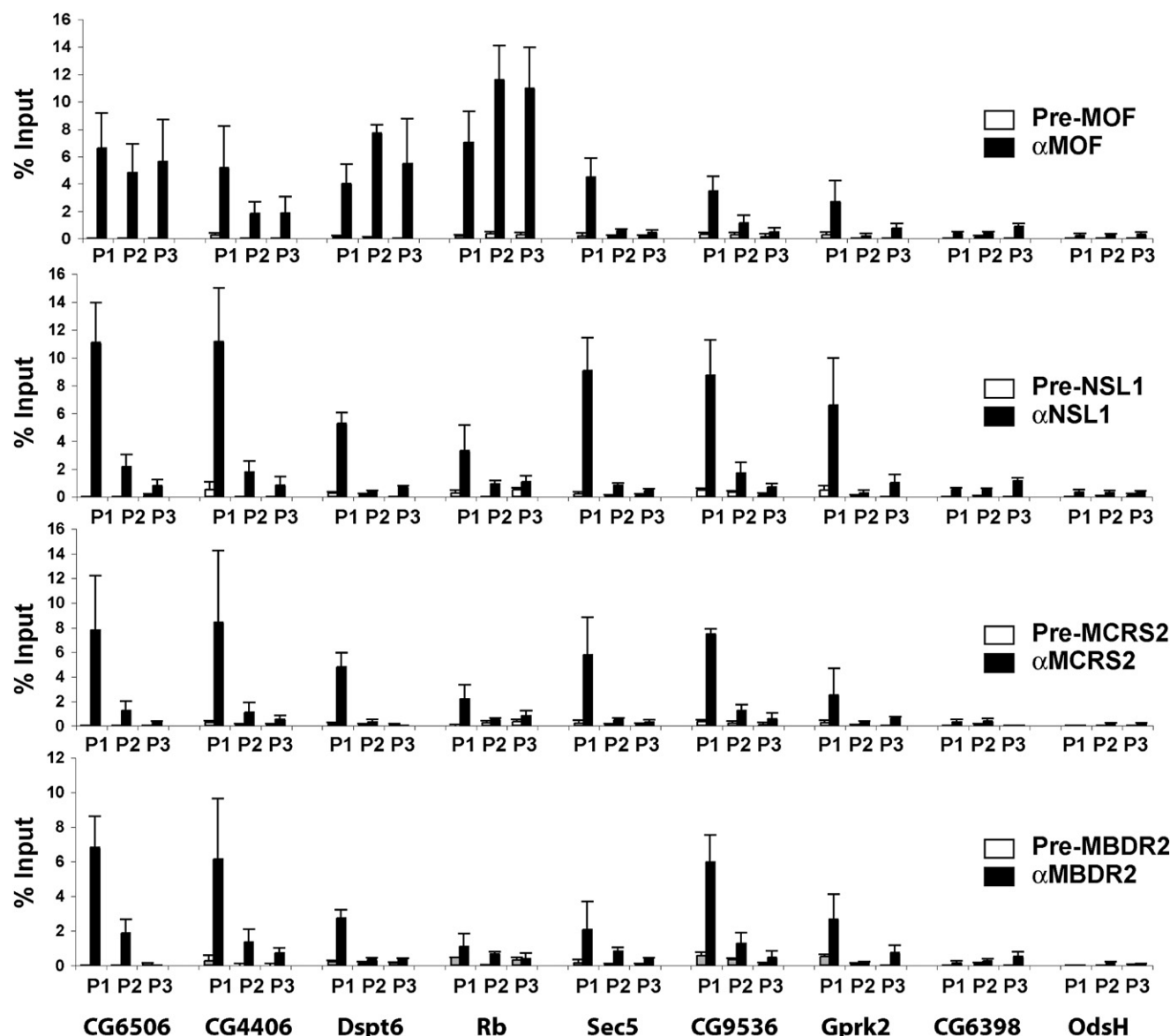
complex members. We further characterized the interaction for the mammalian orthologs hMOF and hNSL1 and could observe that the HAT domain of hMOF (aa 174–458) and the C-terminal domain of hNSL1 (aa 883–1105) were sufficient for interaction. Furthermore, gel filtration as well as multiangle laser light scattering (MALLS) analyses revealed that this interaction was based on a stoichiometry of 1:1 between the two proteins (Figures 1F and 1G; Figure S1C). These results further demonstrate the high degree of evolutionary conservation between NSL1 and MOF interaction across species.

#### The NSL Complex Members Are Localized on MOF Target Promoters

The biochemical association of NSL proteins with MOF prompted us to address whether these proteins are chromatin bound

in vivo. We therefore performed immunofluorescence staining of *Drosophila* third instar larval male and female polytene chromosomes using antibodies against MCRS2, NSL1, MBD-R2, and MOF. All four proteins showed extensive genome-wide colocalization with each other as well as with MOF on interbands of the polytene chromosome squash preparations in both sexes (Figure 2; Figure S2A, available online).

To gain more detailed insight as to where chromatin binding was occurring, ChIP assay was performed on wild-type male salivary glands using antibodies against MCRS2, MBD-R2, NSL1, and MOF; corresponding preimmune sera were used as negative controls. ChIP samples were analyzed using quantitative real-time PCR (qPCR). Seven genes, which were recently identified as MOF target genes either on X chromosomal (CG6506, CG4406, *Dspt6*, and *Rb*) or autosomal (*Sec5*,

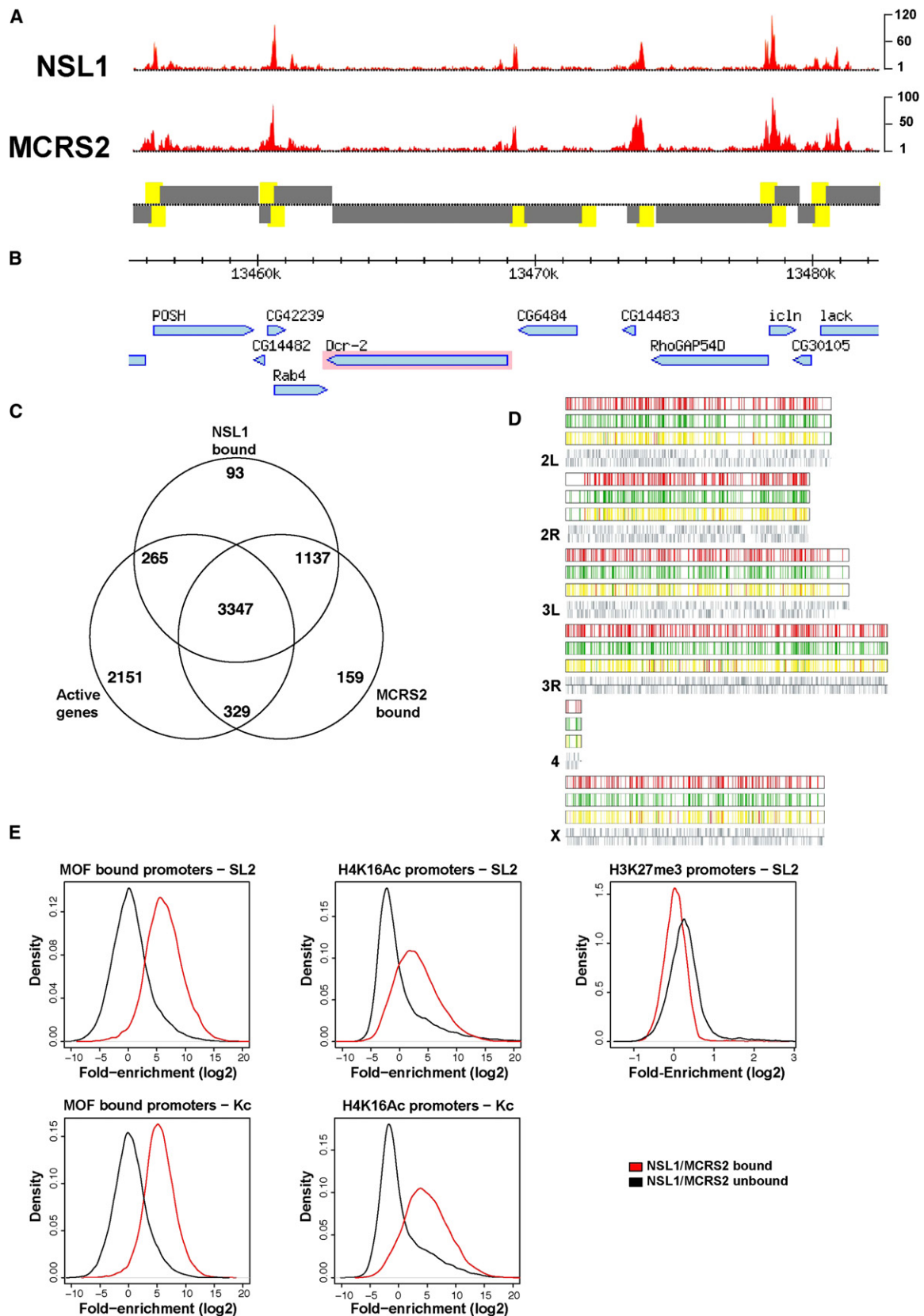


**Figure 3. NSL1, MCRS2, and MBD-R2 Bind to the Promoters of X Chromosomal and Autosomal Genes**

ChIP analysis from larval male salivary glands using antibodies against MOF, NSL1, MCRS2, and MBD-R2. Respective preimmune sera were used as negative controls. Immunoprecipitated DNA was amplified by qPCR with primer sets indicated in the [Supplemental Experimental Procedures](#). Six X-linked genes (CG6506, CG4406, Dspt6, Rb, CG6398, and OdsH) and three autosomal genes (Sec5, CG9536, and Gprk2) were evaluated using primers positioned at the promoter (P1), middle (P2), and end (P3) of the coding sequence. Percentage input is determined as the amount of immunoprecipitated DNA relative to input DNA. Error bars represent standard deviation (StDev) of five independent experiments.

CG9536, and Gprk2) locations (Kind et al., 2008), were tested for NSL binding. Two MOF nontarget genes (CG6398 and OdsH) were used as negative controls. Distribution on each gene was analyzed using primers located toward the promoter (P1), middle (P2), and the end of the coding region (P3) as indicated in Figure 3 (also see [Experimental Procedures](#)). Consistent with previous studies (Kind et al., 2008), MOF showed binding to the promoter regions of autosomal genes, whereas on X chromosomal targets, binding is also distributed along the transcribed regions (compare Figure 3, X targets versus autosomal targets). Interest-

ingly, we found that NSL1, MCRS2, and MBD-R2 are enriched on promoter proximal regions of X chromosomal as well as autosomal MOF target genes (Figure 3). To further address whether this binding pattern is specific to male salivary glands, we performed ChIP from female salivary glands, SL-2 cells, as well as from whole larvae, which showed the same results (Figures S2B–S2D). Furthermore, to get a global perspective, ChIP followed by Solexa deep sequencing analysis revealed that clusters of MCRS2 and NSL1 highly coincide on promoters, suggesting that these proteins act in a complex also in the chromatin





context (Figures 4A, 4B, and 4D; Figure S3A, available online). Over 4000 promoters were bound by NSL1 and MCRS2, ~70% of these corresponding to active genes ( $p < 2.2 \times 10^{-16}$ ; Fisher's exact test; Figure 4C). Correlation with the 5' end of primary transcripts confirmed that NSL1 and MCRS2 binding peaks near transcription start sites. The most frequent distance to the nearest cluster of NSL1 and MCRS2 binding is -45 bp upstream and +2 bp downstream from the transcription start sites, respectively (Figures S3B and S3C). Furthermore, H4K16Ac and MOF binding are enriched at promoters of NSL-bound genes when compared with promoters of genes not bound by MCRS2 and NSL1 (Figure 4E). This happens for both SL-2 and Kc cells ( $p < 2.2 \times 10^{-16}$  for all cases; t test). In contrast, H3K27me3, a mark enriched on inactive genes (Schwartz et al., 2006), is depleted on promoters bound by MCRS2 and NSL1 ( $p < 2.2 \times 10^{-16}$ ; t test). Taken together, the above results reveal the NSL complex as a promoter-bound complex in *Drosophila*.

### NSL1 and MCRS2 Are Required for the Stability of the NSL Complex

*P*-element insertions in *MBD-R2*, *MCRS2*, and *NSL3* lead to early larval lethality in both sexes (see Supplemental Experimental Procedures available online). Therefore, to analyze the functional significance of the NSL complex, we performed tissue-specific RNAi-mediated knockdown in *Drosophila* using the Gal4-upstream activating sequence (UAS) system (Brand and Perrimon, 1993). RNAi-mediated knockdown of MBD-R2, MCRS2, and NSL3 in eye imaginal discs using an *eyeless*-Gal4 driver resulted in a reduced, rough eye phenotype signifying that these proteins are important for development (Figure S4, available online).

To further examine the fate of the NSL complex components upon depletion of individual proteins of the complex, we chose to knockdown MCRS2, MBD-R2, and NSL3 in the salivary glands using *patched*-Gal4,UAS-EGFP/+;UAS-RNAi/+, where EGFP positively marks the knockdown cells. Upon depletion of these proteins, the development of salivary glands is impaired, as observed by reduced gland size (data not shown). Interestingly, depletion of MCRS2 led to a severe reduction of nuclear NSL3 staining, and to a slight reduction in the nuclear staining of MBD-R2 and NSL1 compared to the control glands (Figure S5, available online). Depletion of NSL3 protein resulted in a slight reduction in MCRS2 levels, while the staining of NSL1 and MBD-R2 remained unaffected. In contrast, we did not observe any changes in NSL complex members distribution in MBD-R2 and Z4 depleted glands (Figure S5).

The reduction in staining could be a result of impaired stability of the complex caused by degradation of proteins, changes in

the transcript levels, and/or changes in nuclear and cytoplasmic distribution of the proteins. In order to address this issue, we used Schneider (SL-2) cells to prepare nuclear and cytoplasmic extracts. The efficiency of nuclear and cytoplasmic extract preparations was measured by detecting tubulin levels, which is enriched in the cytoplasm (Figure 5). The effect of the depletion of MCRS2, MBD-R2, and NSL3 could be reproduced in SL-2 cells using the same RNAi sequences as in the flies (Figures 5A and 5B). Western blot analysis for the components of the NSL complex revealed that in MCRS2 depleted cells, NSL3 and NSL2 protein levels were severely reduced, while MBD-R2 and NSL1 protein levels were slightly affected. These effects were specific to the NSL complex members, as MOF and MSL3 levels remained unchanged (Figure 5A). We observed similar results with NSL1 and NSL3 depletion, where depletion of NSL1 severely affected NSL2 while NSL3 depletion affected the levels of NSL3, NSL1, NSL2, and MBD-R2. Levels of other NSL proteins were also moderately compromised in these experiments with the exception of MOF or Tubulin. In contrast, depletion of MOF protein did not affect NSL complex protein levels (Figure 5B).

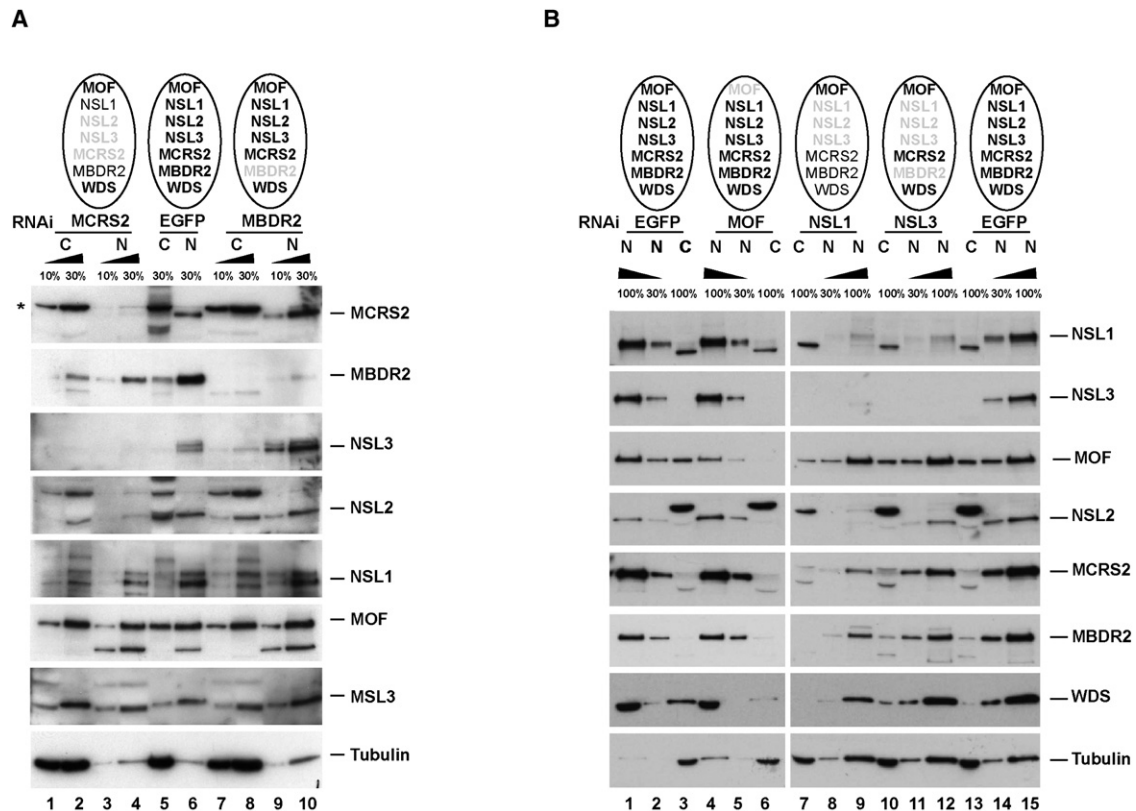
Furthermore, we found that the NSL complex members are bound to their own genes (Figure S6, available online), suggesting autoregulation by the complex. Consistent with these observations we also noticed a reduction in mRNA levels of NSL2 and NSL3 in MCRS2 depleted cells (Figure S7A, available online). Taken together, these results suggest that the integrity of the NSL complex is interdependent.

### Depletion of MCRS2, MBD-R2, and NSL3 Affects the Expression of Target Genes

Polytene chromosome staining as well as the ChIP-Seq approach revealed genome-wide association of NSLs to the promoter regions of a large number of target genes. One of the most obvious questions that arise is whether binding of the NSL complex to target genes is implicated in the regulation of gene expression. *P*-element insertion mutants of MCRS2, MBD-R2, and NSL3 are early larval lethal. So, to measure the effect on transcriptional activity, total RNA was isolated from third instar larval MCRS2, NSL3, and MBD-R2 depleted salivary glands and processed for quantitative reverse transcription (qRT)-PCR analysis (see Experimental Procedures). We chose a set of MCRS2 and NSL1 target genes on both the X chromosome and autosomes to test for a change in transcript levels upon MCRS2, NSL3, and MBD-R2 knockdown. Binding of NSL1, MCRS2, MBD-R2, and MOF to the promoters of these genes was confirmed by ChIP (Figure 3; Figures S2B–S2D). Strikingly, we found a severe reduction in transcript levels for most of the tested NSL complex target genes upon depletion of MCRS2, NSL3, or MBD-R2, respectively (Figures 6B–6D).

### Figure 4. ChIP-Seq Analysis Reveals that NSL1 and MCRS2 Bind to Promoters Genome Wide

- (A) Snapshot of a chromosome 2R region obtained using Genomatix EIDorado; tracks show read counts for NSL1 and MCRS2 (red). Gene promoters (yellow blocks) and transcripts (gray blocks) are highlighted.
- (B) Flybase genes located in the selected region.
- (C) Overlap between MCRS2-bound, NSL1-bound, and actively expressed genes in salivary glands.
- (D) Chromosomal maps of genes bound by MCRS2 (red) and NSL1 (green). The extent of overlap is shown in yellow. For each chromosome genes on the + strand (top) and - strand (bottom) are shown in gray.
- (E) Distribution of fold enrichments for MOF binding, H4K16ac, and H3K27me3 in SL-2 and Kc cells for genes bound (red) and unbound (black) by MCRS2 and NSL1 ( $p < 2.2 \times 10^{-16}$  in all cases; see Supplemental Experimental Procedures).



**Figure 5. The Stability of the NSL Complex Is Compromised upon Depletion of NSL1, NSL3, and MCRS2**

(A) Western blot analysis of the NSL complex members in MCRS2-RNAi, EGFP-RNAi, and MBD-R2-RNAi treated SL-2 cells.

(B) Western blot analysis of the NSL complex members in EGFP-RNAi, MOF-RNAi, NSL1-RNAi, and NSL3-RNAi treated SL-2 cells. C and N are the lanes loaded with cytoplasmic and nuclear extract, respectively. Percentage of extract loaded is also shown. Asterisks indicates the position of a cross-reacting band recognized by MCRS2 in the cytoplasmic fraction. Schematic representation of the NSL protein levels is shown on top of each RNAi condition, where gray color indicates reduced protein levels for the respective protein.

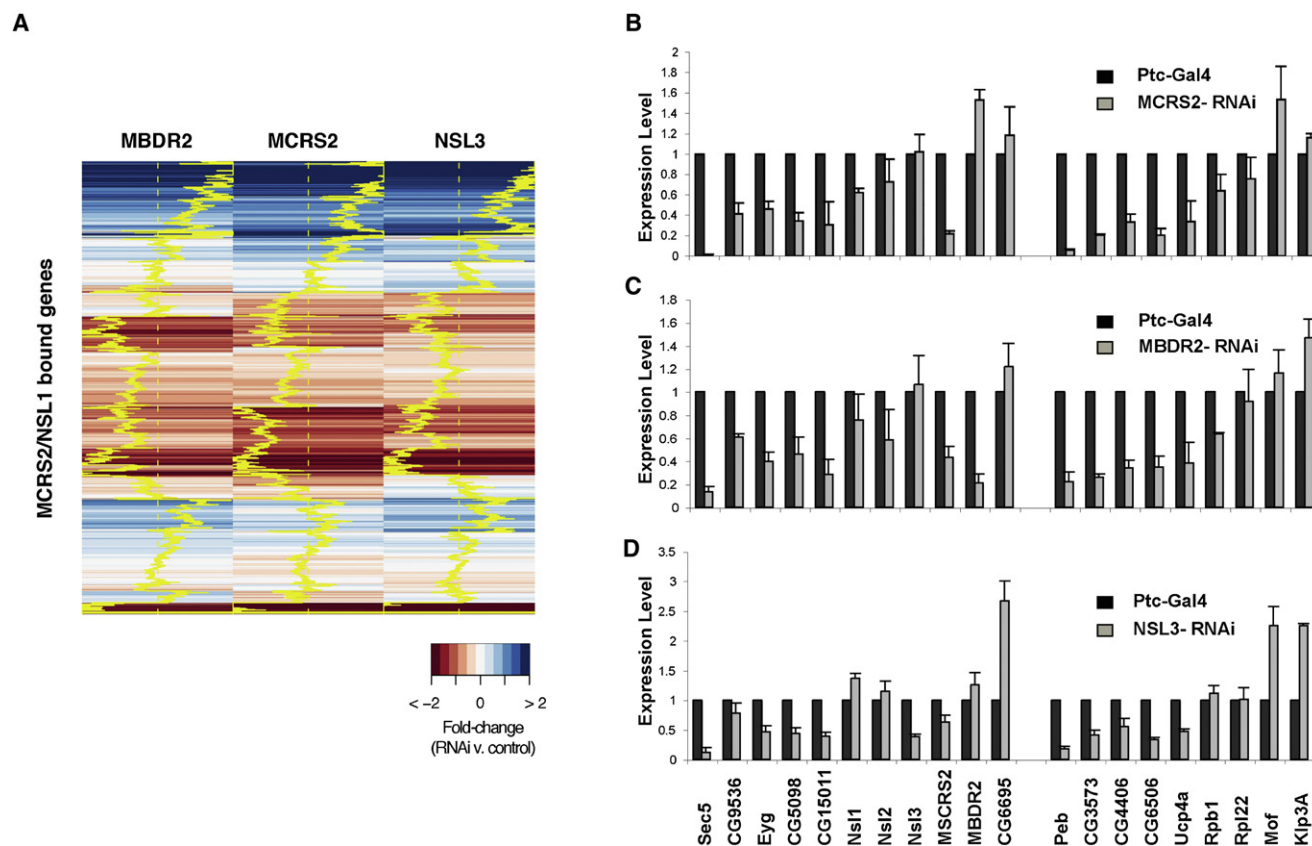
Similar results were obtained when SL-2 cells were depleted of MCRS2 (Figure S7B).

To address whether these effects were also observed on a genome-wide scale we performed gene expression profiling of RNA isolated from MCRS2, MBD-R2, and NSL3 depleted salivary glands using Affymetrix arrays (Figure 6A). We observed that RNAi depletion of any of these proteins results in a very similar pattern of gene expression changes (Figure 6A; Pearson correlation  $> 0.86$  for all comparisons). A total of 5045 genes were differentially expressed in the MCRS2 depleted glands, 3996 genes in the MBD-R2 depleted glands and 4213 genes in the NSL3 depleted glands. In all cases, there was a significant enrichment of differentially expressed genes among the bound ones ( $p < 2.2e-16$  for all three RNAi experiments; Fisher's exact test). Moreover, correlation of genome-wide promoter binding with the expressed genes in the salivary glands revealed 3347 out of 6092 active gene promoters to be bound by both NSL1 and MCRS2 (Figure 4C).

We found similar numbers of up- and downregulated genes among NSL targets, indicating that the loss of the complex is associated with both gene activation and repression. These results were confirmed by qRT-PCR (Figure 6B). A notable

exception was MCRS2: The qRT-PCR validation strongly suggested a major activating role. Such a role is difficult to observe using standard array procedures that usually assume a moderated and balanced change on gene expression upon the studied perturbation. To account for this, we performed linear regression between the qRT-PCR and the microarray-generated data (see Supplemental Experimental Procedures). The results indicated a global downregulation upon MCRS2 RNAi depletion that was not observed for any of the other RNAi treatments (data not shown).

Gene expression analysis was also done using salivary glands of male larva that carry the *mof*<sup>2</sup> mutation, where a premature stop codon prevents formation of a functional MOF protein (Hilfiker et al., 1997). The effect of NSL depletion on gene expression was generally more severe in its magnitude than the one observed in the absence of MOF (compare Figure 6 with Figure S7C). Furthermore, MCRS2, NSL3, and MBD-R2 depletion did not show a bias for X chromosomal genes, with genes on all chromosomes being similarly affected. Interestingly, binding of the NSL complex members MCRS2, MBD-R2, and NSL1 to chromatin was unaffected in *mof*<sup>2</sup> mutant salivary glands as shown by polytene chromosome staining (Figure S8, available



**Figure 6. Depletion of MCRS2, MBD-R2, and NSL3 Affects Gene Expression on X Chromosomes and Autosomes**

(A) Heatmap displaying gene expression changes upon RNAi depletion of MCRS2, MBD-R2, and NSL3 (columns) for genes bound by MCRS2 and NSL1 (rows). Downregulated genes are shown in red. Upregulated genes are shown in blue. Trace lines (yellow) centered in the middle of each column represent the magnitude of the change according to the scale. Both rows and columns were clustered using hierarchical clustering.

(B–D) Quantitative RT-PCR analysis of *Sec5*, *CG9536*, *Eyg*, *CG5098*, *CG15011*, *NSL1*, *NSL2*, *NSL3*, *MCRS2*, *MBD-R2*, and *CG6695* located on autosomes and *Peb*, *CG3573*, *CG4406*, *CG6506*, *Ucp4a*, *Rpb1*, *Rpl22*, *Mof*, and *Klp3A* located on the X chromosome. Wild-type control *Ptc-GAL4* salivary gland expression (black bars) and RNAi depleted salivary gland expression (gray bars). *Ptc-Gal4;UAS-MCRS2-RNAi* (B), *Ptc-Gal4;UAS-MBD-R2-RNAi* (C), and *Ptc-Gal4;UAS-NSL3-RNAi* (D). Error bars represent standard deviation (StDev) of three independent experiments. Expression levels were normalized against the respective genomic DNA.

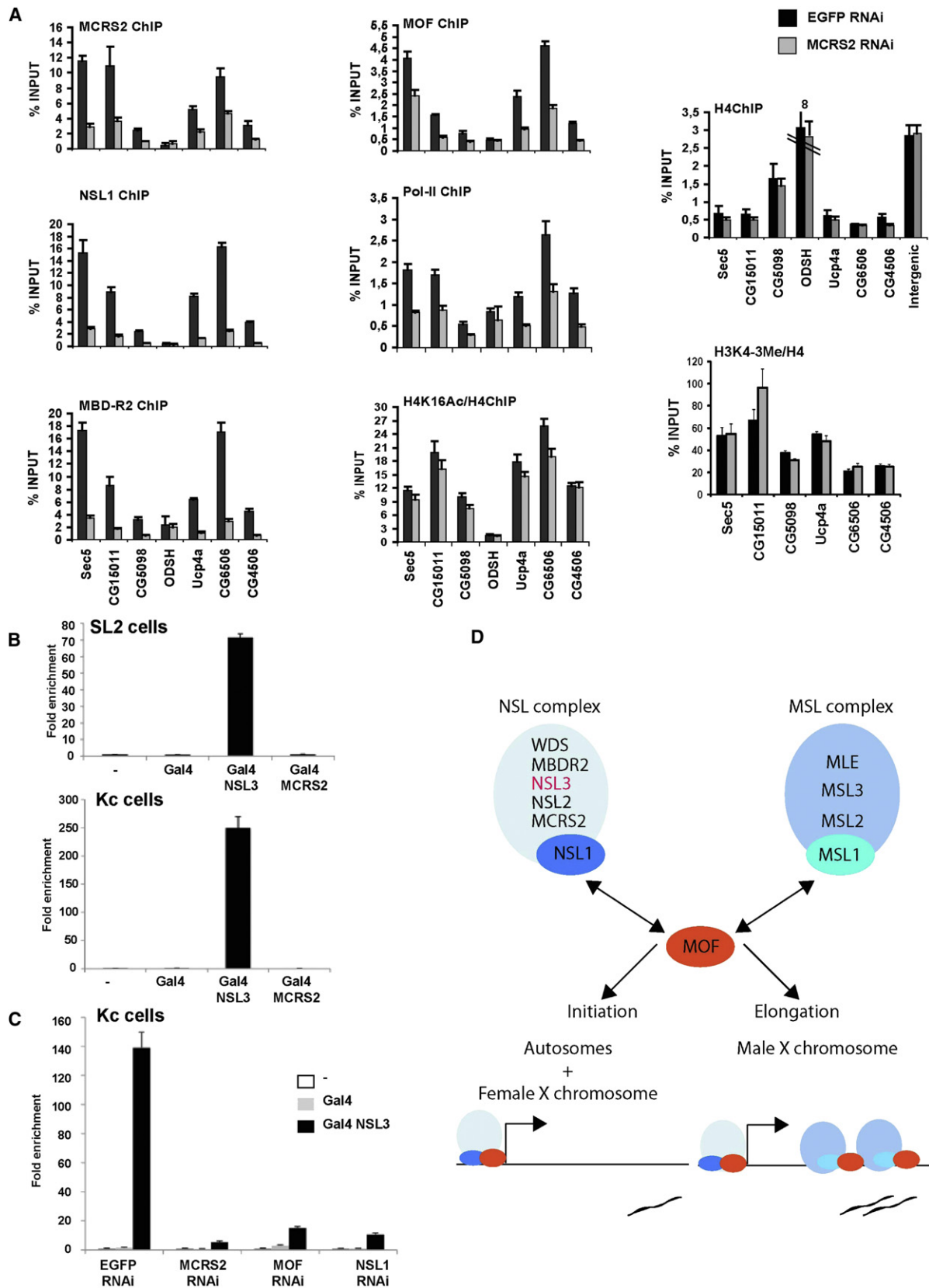
online). These results suggest a role of NSLs in transcription regulation upstream of MOF function. We therefore propose the evolutionarily conserved NSL complex as an important transcriptional regulator in *Drosophila*.

#### Depletion of MCRS2 Affects the Integrity of the NSL Complex at Promoters

To further assess whether MCRS2 is required to stably maintain the integrity of the NSL complex on chromatin, we performed ChIP for NSL1, MBD-R2, and MOF on NSL1 and MCRS2 target genes upon MCRS2 depletion. We chose RNAi-mediated knockdown in SL-2 cells, since salivary glands are greatly reduced in size upon MCRS2 knockdown, thus limiting the amount of sample available for ChIP. As expected, depletion of MCRS2 led to reduced levels of MCRS2 at target promoters (Figure 7A) and up to 90% of knockdown efficiency (Figure 5A). Interestingly, depletion of MCRS2 resulted in a severe reduction in the binding of NSL1 and MBD-R2 to these target genes

(Figure 7A), while their overall protein levels were only moderately affected (Figure 5). These results demonstrate that MCRS2 is required to stably maintain the integrity of the NSL complex at target promoters.

Furthermore, although the protein level of MOF was unaffected upon depletion of MCRS2 (Figure 5A), MOF binding to target promoters was substantially reduced (Figure 7A). Similar results were obtained by performing ChIP on whole larvae using an MCRS2 mutant (Figures S9A–S9C, available online). We also tested the effects of MCRS2 depletion on the levels of H4K16Ac and H3K4me3, as both marks are associated with active genes. ChIP using antibody against histone H4 was used as control. We observed no significant reduction in H3K4me3 levels at these promoters, however, acetylation at H4K16 was moderately reduced upon depletion of MCRS2 (Figure 7A). Moderate reduction of H4K16Ac may be explained by the dynamic nature of MOF interaction at these sites or a low H4K16Ac turnover. However, another very likely possibility could be that since these





ChIPs were performed in SL-2 cells, presence of the MSL complex may still be responsible for overall acetylation levels. Indeed, when comparing MOF levels at the 5' versus 3' end of X-linked genes, we observed a more significant reduction at the promoters versus the 3' end upon MCERS2 depletion (Figure S9D). Consistent with these observations, when we repeated the ChIP in Kc cells depleted of MCERS2, where the MSL complex is not assembled, NSL1, MBD-R2, and MOF as well as H4K16 levels were severely affected (Figure S10A, available online).

As mentioned earlier, upon depletion of MCERS2, MBD-R2, and NSL3, the mRNA levels of most of the tested target genes were reduced. We were next interested to know whether the reduction in expression of these target genes is a result of defects in the recruitment of RNA polymerase II to their promoters. Depletion of MCERS2 led to a severe loss of NSL1, NSL2, NSL3, and MBD-R2 from gene promoters. We therefore chose to assay the binding of RPB3—a subunit of the core RNA polymerase II (Adelman et al., 2005) on the target genes in MCERS2 depleted cells. Interestingly, the levels of RPB3 on target promoters were significantly reduced in MCERS2 depleted cells (Figure 7A). However, overall nuclear RNA polymerase II levels remain unaffected as shown by western blot (Figure S10B). These results suggest that the reduction in expression upon loss of the NSL complex from its target genes might be a consequence of altered recruitment of RNA polymerase II to the promoters, suggesting a function of NSLs in early transcriptional regulation.

### NSL3-Driven Transcription Activation Is Modulated by NSL1, MCERS2, and MOF

Next, we were interested in studying whether the NSL complex has the potential to activate transcription. We therefore expressed two components of the NSL complex, NSL3 and MCERS2, fused to the Gal4 DNA-binding domain in SL-2 or Kc cells cotransfected with a UAS-driven luciferase reporter. Interestingly, expression of Gal4-NSL3 resulted in potent activation of the reporter (Figure 7B). In contrast, transfection of Gal4 alone or Gal4-MCERS2 did not show activation of the reporter gene. We next tested whether activation by Gal4-NSL3 is sensitive to the endogenous levels of NSL complex members. For this purpose we depleted NSL1 and MCERS2 as well as MOF in cells trans-

fected with Gal4-NSL3 or Gal4-DBD as a control. Reduction in levels of NSL1 or MCERS2 severely affected the transcription activation mediated by NSL3. Similarly, depletion of MOF substantially reduced reporter activation (Figure 7C). These results strongly suggest that the NSL complex members and MOF work synergistically to activate transcription.

## DISCUSSION

Here we report the biochemical purification and functional analysis of the NSL complex. The members of this complex are evolutionarily conserved between fruit flies and humans (Mendjan et al., 2006). Intriguingly, we find that 55% of all active genes are bound by the NSL complex in *Drosophila*. We show that loss of binding of these factors severely affects expression of target genes. Interestingly, loss of NSL proteins affects MOF binding as well as RNA pol II recruitment to promoters. Furthermore, we show that the NSL complex and MOF work synergistically to activate transcription. Taken together, these results identify the NSL complex as a transcriptional regulator in the *Drosophila* genome.

### Genome-wide Colocalization of the NSL Complex Members on Gene Promoters

We present several lines of evidence that the NSL complex resides at gene promoters. First, immunostaining of polytene chromosomes with antibodies against NSL complex members and MOF highly colocalizes in both sexes (Figure 2; Figure S2A). Second, NSL1, MCERS2, MBD-R2, and MOF colocalize at the promoters of the tested MOF target genes in salivary glands as shown by ChIP experiments (Figure 3). The NSL complex binding at promoters is not restricted to salivary glands, as we found essentially the same results using chromatin from SL-2 cells and whole larvae (Figures S2B–S2D). Finally, using ChIP and Solexa sequencing, we show that NSL1 and MCERS2 signals highly overlap at the promoters of ~4000 genes in salivary glands (Figure 4). There is a striking correlation between promoter-bound NSL complex members and MOF as over 70% of previously identified MOF-bound promoters in SL-2 and Kc cells (Kind et al., 2008) overlap with promoter-bound NSL1 and MCERS2 in male salivary glands (Figure 4E).

### Figure 7. Depletion of MCERS2 Affects the Integrity of the NSL Complex at Target Promoters

(A) ChIP using antibodies against MCERS2, NSL1, MBD-R2, MOF, Rpb3 (Pol II), Histone H4, and H4K16Ac on four X-linked genes (*Odsh*, *Ucp4A*, *CG6506*, and *CG4406*) and three autosomal genes (*Sec5*, *CG15011*, and *CG5098*) in EGFP-RNAi treated (black bars) and MCERS2-RNAi treated (gray bars) SL-2 cells. Primers were positioned at the promoters of the mentioned genes. Exact positions of the primers are described in the Supplemental Data. Percentage input is determined as the amount of immunoprecipitated DNA relative to input DNA. Error bars represent standard deviation (StDev) of four independent experiments.

(B) Tethering NSL3 via the GAL4 DNA-binding domain activates transcription of the luciferase reporter under the UAS promoter in SL-2 cells (top) and Kc cells (bottom). Transient transfection of either empty vector (–), vector expressing the GAL4 DNA-binding domain (Gal4), GAL4NSL1, GAL4NSL3, or GAL4MCERS2. Error bars represent standard deviation (StDev) of three independent experiments.

(C) NSL3-mediated transcription activation is affected by depletion of MCERS2, NSL1, and MOF. Transient transfection of either empty vector (–), vector expressing GAL4 DNA-binding domain (Gal4), or GAL4NSL3 in cells treated with EGFP RNAi, MCERS2 RNAi, MOF RNAi, or NSL1 RNAi. y axis shows fold enrichment compared to vector control. Error bars represent standard deviation (StDev) of three independent experiments.

(D) Summary model: MOF protein is present in two distinct complexes. The classical MSL complex (or the dosage compensation complex) that is enriched on the 3' end of X-linked genes and the NSL complex that is a promoter-bound complex. Within these two complexes MOF directly interacts with either MSL1 or NSL1, two PEHE domain-containing proteins. The stability of the NSL complex is interdependent, MCERS2, NSL1, and NSL3 playing a major role in the overall stability of the complex. Tethering NSL3 via Gal4 DNA-binding domain to a UAS containing heterologous reporter activates transcription in both SL-2 and Kc cells and this activation is modulated by the presence of NSL1, MCERS2, or MOF. We propose a working model that synergistic interaction of the NSL complex with MOF at promoters might promote transcription initiation, while interaction of MOF with the MSL complex at the 3' end of X-linked genes may facilitate transcription elongation.

It was previously shown that depletion of MSL1 led to the loss of MOF from the 3' end of X-linked target genes in males, but, surprisingly, MOF binding at the promoters of the same genes was unaffected (Kind et al., 2008). The results shown in the present study unravel that MOF binds to the promoters in an association with the NSL complex and that this association is not restricted to the male X chromosome, but is genome wide in both sexes. The observation that MOF binding at promoters is reduced upon depletion of MCSR2 clearly suggests that the NSL complex contributes to the recruitment of MOF to target gene promoters (Figure 7). However, chromatin binding of the NSL complex members is not affected in the absence of MOF, implicating NSLs in transcription regulation upstream of MOF. These data reveal a hierarchy of recruitment at these target promoters. Furthermore, it is noteworthy that NSL mutants are early larval lethal in both sexes, whereas, Mof mutants are male lethal due to dosage compensation defects, but females are viable to the adult stage albeit being sterile (Gelbart et al., 2009). These observations suggest that NSL1, NSL2, NSL3, MCSR2, and MBD-R2 are important regulators during fly development and play additional functions very likely independent of MOF.

#### **MCSR2 Is Required to Stably Maintain the NSL Complex on Promoters**

We find that the integrity of the NSL complex is interdependent, very likely operating at both transcriptional and posttranscriptional levels. Our study identifies NSL1 and MCSR2 among the key players in maintaining the integrity of the NSL complex. Depletion of MCSR2 not only affected protein levels of NSL2 and NSL3, but ChIP analysis also revealed that in the absence of MCSR2, the rest of the complex is unable to achieve efficient chromatin targeting. There is an interesting resemblance of the above observations with that of the MSL complex where the MSL2 protein is required for complex stability (Beckmann et al., 2005; Gebauer et al., 2003; Grskovic et al., 2003; Kelley et al., 1997).

MCSR2 contains a FHA domain at its C terminus. The FHA domain is a phosphopeptide recognition domain found in many regulatory proteins. It displays specificity for phosphothreonine-containing epitopes but also recognizes phosphotyrosine with relatively high affinity (Durocher and Jackson, 2002). It is provocative to speculate that MCSR2 could recognize the phosphorylated forms of the members of the NSL complex through the FHA domain to stabilize and recruit the complex to its target promoters.

#### **The NSL Complex Regulates Transcription at Promoters**

Depletion of NSL components resulted in expression changes of many of the NSL-bound genes on autosomes as well as on the X chromosome. Importantly, magnitude and extent of the expression effect was more severe than the one observed in the absence of MOF, suggesting a MOF-independent mechanism for NSL function. However, it remains possible that the role of MOF at target promoters is to fine tune gene expression by modulating NSL function. This is further supported by our observations that NSL3-mediated transcription activation is sensitive to endogenous MOF levels. The broad polytene chromosome

staining as well as ChIP-Seq analysis of the NSLs shows that many chromatin regions are occupied by the NSL complex, which is likely to be involved in the regulation of a wide spectrum of genes (Figure 7D).

The observation that RNA polymerase II recruitment to target genes is affected upon depletion of MCSR2 suggests that the NSL complex is involved in early steps of transcription. Taking into account that over 4000 promoters are bound by both NSL1 and MCSR2, correlating with ~55% of the active genes, and that the expression of most of the NSL target genes tested in this study is compromised upon the loss of NSL complex members in salivary glands, several fundamental questions emerge: How is the NSL complex targeted to responsive promoters? Which step of transcription initiation requires the presence of the NSL complex?

The NSL complex could possibly recognize specific DNA sequences at promoter regions or is brought to promoters through the interactions with components of the transcription machinery or regulatory proteins present on the promoters. These interactions would most probably be transient, and reflect an inherent instability, as no such proteins were found purifying with the NSL complex members.

Another very likely possibility is that the complex recognizes histone marks. The NSL proteins harbor a rich composition of chromatin-binding domains. For example, WDS is a protein with the potential of binding to histone H3. It belongs to the WD family and consists of seven WD40 repeats (Hollmann et al., 2002). It is known that these repeats are involved in protein-protein interaction and are present in many chromatin-associated complexes (Cao et al., 2002). The mammalian ortholog of this protein, WDR5, binds specifically to the N-terminal tail of histone H3 and thereby helps to recruit the MLL containing H3K4-specific methyltransferase complex to its target promoters (Wysocka et al., 2005). Methylated H3K4 is coupled to transcription activation (Zhang and Reinberg, 2001). We show that H3K4me3 is unaffected in the absence of MCSR2, indicating that the NSL complex recruitment could be downstream of this activation mark. WDS could potentially have analogous functions in recruiting the NSL complex to target promoters in order to regulate the transcription of target genes. Another member of the complex is MBD-R2, which has a Tudor and methyl binding (MBD) domains (Taipale et al., 2005). Tudor domains share similarity to chromodomains, which also bind methylated residues (Lachner et al., 2001). In yeast, Tudor domains have been shown to bind methylated H3K79 (Huyen et al., 2004). MBD domains bind to methylated DNA and are involved in transcriptional repression in mammals (Bird, 2002). However, DNA methylation happens much more seldom in *Drosophila* than in mammals and its function remains unclear (Lyko et al., 2000). MBD-R2 is therefore another candidate that could recognize modified histones and target the NSL complex to promoters in *Drosophila*.

It is important to emphasize that we have shown earlier that the mammalian orthologs of the *Drosophila* NSL complex members were found copurifying with human MOF (Mendjan et al., 2006). These observations have been further confirmed recently (Cai et al., 2010). The mammalian NSL complex was shown to have a more relaxed substrate specificity acetylating histone H4 lysine 5, 8, and 16, in comparison to the MSL complex where MOF

preferentially acetylates histone H4 lysine 16 (Cai et al., 2010; Mendjan et al., 2006). Similar to the *Drosophila* MOF protein which can acetylate nonhistone substrates (Buscaino et al., 2003; Morales et al., 2004), mammalian MOF alone or in association with MSL1v1 (NSL1) has recently been shown to acetylate other nonhistone substrates such as TIP5 or p53, respectively (Li et al., 2009; Zhou et al., 2009). It therefore remains to be seen in the future whether the mammalian NSL complex could also play a more global role in transcriptional regulation. Besides transcription, NSL components have also recently been implicated in other cellular processes (Dobbelaere et al., 2008; Nybakken et al., 2005). In summary, the biochemical purification and functional characterization of NSL proteins described in this study reveals an important evolutionarily conserved genome-wide promoter-bound complex that acts as a major transcriptional regulator in *Drosophila*.

## EXPERIMENTAL PROCEDURES

### Biochemical Purifications

Biochemical purifications from SL-2 cells expressing tagged proteins (TAP-MCRS2 or HA/FLAG NSL1) were always accompanied with control purifications from wild-type SL-2 nuclear extract to be able to compare specific enrichment.

TAP purifications were done essentially as described (Rigaut et al., 1999) with the following changes in the protocol: All buffers contained 25 mM HEPES (pH 7.6) instead of Tris-HCL, KCl instead of NaCl, 1/100 volume of RNasin (Promega), 0.2% Tween-20, and 20% glycerol. For details on HA/FLAG purifications see Supplemental Experimental Procedures.

### Silver Staining and Mass Spectrometry

Silver staining and mass spectrometry were performed as previously described (Shevchenko et al., 1996). Purified protein samples were prepared for mass spectrometry in two ways. Either individual silver-stained protein bands were digested in gel with trypsin (Figure 1A) or complex elutions were electrophoresed for a few minutes, nonseparated proteins stained with Coomassie blue, total band excised, and trypsin digested (Table S1). The samples were separated for 45 min on a nano-flow 1D-plus Eksigent (Eksigent, Dublin, CA) HPLC system coupled to a qStar Pulsar i quadrupole time-of-flight mass spectrometer (Applied Biosystems, Darmstadt, Germany). Analysis was performed as described (Fraterman et al., 2007). Taxonomy parameter was restricted to *Drosophila*, and peptides below a score of 18 were excluded. All proteins were identified with a summed peptide MASCOT score above 45 in at least two independent experiments.

### Biochemical Interaction Assays

Recombinant full-length NSL1, MCRS2, and MOF were either singularly or coexpressed using the baculovirus coexpression system. Details of the purification are available upon request.

For the hMOF/NSL1 interaction, a His-tag fusion of the C-terminal fragment of hNSL1 (883–1105) and untagged MOF HAT domain (174–458) were coexpressed in *E. coli* BL21Star(DE3) (Invitrogen) from pProEXHTb (Invitrogen) and pRSFDuet-1 (Novagen) expression vectors, respectively. The complex was first purified by affinity chromatography using Ni<sup>2+</sup> resin. In parallel, a His-tag version of hMOF(174–458) was expressed from pProEXHTb. The His tag was removed by TEV protease following the Ni<sup>2+</sup> affinity chromatography. The gel-filtration profiles were obtained using Superdex 200 column (GE Healthcare).

### RNAi in SL-2 Cells, Nuclear, and Cytoplasmic Fractions

RNAi of SL-2 and Kc cells was performed as described in Worby et al. (2001) with the following modifications. All knockdown cells were transfected with 50 µg dsRNA using Lipofectamine (Invitrogen). The cells were harvested after 6 days for MCRS2 RNAi. EGFP control RNAi experiments were performed in

parallel. Nuclear and cytoplasmic extracts were done from 2 × 10<sup>6</sup> cells. The cell pellet was dissolved in HEMG 40 (HEMG buffer contains 25 mM HEPES pH 7.6, 0.5 mM EDTA, 5 mM MgCl<sub>2</sub>, and 20% glycerol; so HEMG 40 is HEMG with 40 mM KCl), and placed for 10 min on ice for swelling. NP40 was added to 1% final concentration, vortexed for 7 s, and centrifuged at 3000 rpm for 5 min. The supernatant containing cytoplasmic extract was carefully removed to a new eppendorf tube and 4× Laemmli buffer was added. The nuclei pellet was washed extensively with HEMG 150 and lysed in 2× Laemmli buffer.

### RNAi Knockdown in Larval Salivary Glands

The homozygous transgenic RNAi fly strains carrying *UAS-MCRS2-RNAi*, *UAS-NSL3-RNAi*, *UAS-MBD-R2-RNAi*, and *UAS-Z4-RNAi* were independently crossed to *Ptc-Gal4*, *UAS-eGFP*, and raised at 25°C to obtain efficient knockdown. Salivary glands from third instar larvae were dissected in PBS and processed for either immunofluorescence antibody staining or for RNA isolation and qRT-PCR analysis. Antibodies against NSL1, NSL3, MCRS2, MBD-R2, MOF, H4K16Ac, MSL1, and Z4 for IF were used at 1:100 dilution.

### RNA Isolation and Quantitative PCR Analysis

For qRT-PCR, RNA and corresponding genomic DNA from SL-2 cells or from salivary glands were isolated simultaneously using the AllPrep DNA/RNA mini kit (QIAGEN), the RNA column was DNaseI treated, and 300 ng of total RNA was used in a RT reaction. qRT-PCR was performed on a Applied Biosystems (AP) Cyclex7500 with SYBR detection, and the amplification curves were analyzed with the corresponding AP software. Each qRT-PCR was repeated at least three times, values were normalized to corresponding genomic DNA values, and the standard deviation within each experiment was calculated.

qPCR analysis of the ChIP samples was performed using the SYBR Green PCR master mix (Applied Biosystem), 100 ng of each forward and reverse primer, and 1 µl immunoprecipitated DNA, in an ABI7500 real-time PCR thermocycler (Applied Biosystems, Inc.). The formula [%ChIP/input] =  $\frac{[E^{Ct_{input}} - Ct_{ChIP}]}{[E^{Ct_{input}} - Ct_{ChIP}]} \times 100\%$  (*E* represents primer annealing efficiency) was used to calculate the percentage of DNA recovery after ChIP, as compared to the amount of input material. The primers designed for the promoter (P1), middle (P2) and the end (P3) of the gene were used for the binding analysis.

### Immunofluorescence and Confocal Microscopy

Preparation of polytene chromosomes was performed as described (<http://www.igh.cnrs.fr/equip/cavalli/Lab%20Protocols/Immunostaining.pdf>). MOF and MSL1 antibodies were used at 1:500 dilution, MBD-R2 at 1:300, and MCRS2 and NSL1 at 1:50. Immunofluorescence stainings of whole-mount salivary glands were essentially done as described (Beuchle et al., 2001). Images were captured with AxioCamHR CCD camera on a Leica SP5 (Leica Microsystems) using an Apochromat NA 1.32 oil immersion objective. Images were arranged with Adobe Illustrator.

### ACCESSION NUMBERS

Microarray expression data are available in the ArrayExpress database under accession number E-MEXP-2652, and the ChIP-seq data are available under accession number E-MTAB-214.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, 10 figures, and 2 tables and can be found with this article online at doi:10.1016/j.molcel.2010.05.021.

### ACKNOWLEDGMENTS

We are grateful to the members of the lab for useful discussions and critical reading of the manuscript. We thank Nic Tapon and Ditte Anderson for communication of results prior to publication. We thank Matthias Wilm and Marc Gentzel for support for mass spectrometry analysis. We thank Vladimir Benes, Tomi Baehr-Ivacevic, and Johanthan Blake in the Genomics core

facility for processing of the ChIP-sequencing samples. We thank John T. Lis for providing us the anti-RPB3 antibody. We are grateful to Stephen Cusack for support. We thank Marc Jamin for help with the MALLS experiment. We thank Erinc Hallacali and Ibrahim Ilik for critical reading of the manuscript. This work was funded from Deutsche Forschungsgemeinschaft-Sonderforschungsbereiche (Transregio) grant and Epigenome Network of Excellence under EU Framework Programme 6 to A.A. P.G. was supported by an "E-STAR" fellowship funded by the EU's FP6 Marie Curie Host fellowship for Early Stage Research Training under contract number MEST-CT-2004-504640.

Received: September 2, 2009

Revised: January 27, 2010

Accepted: April 6, 2010

Published: June 24, 2010

## REFERENCES

- Adelman, K., Marr, M.T., Werner, J., Saunders, A., Ni, Z., Andrulis, E.D., and Lis, J.T. (2005). Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol. Cell* 17, 103–112.
- Akhtar, A., and Becker, P.B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyl transferase essential for dosage compensation in *Drosophila*. *Mol. Cell* 5, 367–375.
- Alekseyenko, A.A., Larschan, E., Lai, W.R., Park, P.J., and Kuroda, M.I. (2006). High-resolution ChIP-chip analysis reveals that the *Drosophila* MSL complex selectively identifies active genes on the male X chromosome. *Genes Dev.* 20, 848–857.
- Alekseyenko, A.A., Peng, S., Larschan, E., Gorchakov, A.A., Lee, O.K., Kharchenko, P., McGrath, S.D., Wang, C.I., Mardis, E.R., Park, P.J., et al. (2008). A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. *Cell* 134, 599–609.
- Beckmann, K., Grskovic, M., Gebauer, F., and Hentze, M.W. (2005). A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in *Drosophila*. *Cell* 122, 529–540.
- Beuchle, D., Struhl, G., and Muller, J. (2001). Polycomb group proteins and heritable silencing of *Drosophila* Hox genes. *Development* 128, 993–1004.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21.
- Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118, 401–415.
- Buscaino, A., Kocher, T., Kind, J.H., Holz, H., Taipale, M., Wagner, K., Wilm, M., and Akhtar, A. (2003). MOF-regulated acetylation of MSL-3 in the *Drosophila* dosage compensation complex. *Mol. Cell* 11, 1265–1277.
- Cai, Y., Jin, J., Swanson, S.K., Cole, M.D., Choi, S.H., Florens, L., Washburn, M.P., Conaway, J.W., and Conaway, R.C. (2010). Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. *J. Biol. Chem.* 285, 4268–4272.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298, 1039–1043.
- Dobbelaere, J., Josue, F., Suijkerbuijk, S., Baum, B., Tapon, N., and Raff, J. (2008). A genome-wide RNAi screen to dissect centriole duplication and centrosome maturation in *Drosophila*. *PLoS Biol.* 6, e224.
- Durocher, D., and Jackson, S.P. (2002). The FHA domain. *FEBS Lett.* 513, 58–66.
- Fraterman, S., Zeiger, U., Khurana, T.S., Wilm, M., and Rubinstein, N.A. (2007). Quantitative proteomics profiling of sarcomere associated proteins in limb and extraocular muscle allotypes. *Mol. Cell. Proteomics* 6, 728–737.
- Gebauer, F., Grskovic, M., and Hentze, M.W. (2003). *Drosophila* sex-lethal inhibits the stable association of the 40S ribosomal subunit with msl-2 mRNA. *Mol. Cell* 11, 1397–1404.
- Gelbart, M.E., Larschan, E., Peng, S., Park, P.J., and Kuroda, M.I. (2009). *Drosophila* MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nat. Struct. Mol. Biol.* 16, 825–832.
- Gilfillan, G.D., Straub, T., de Wit, E., Greil, F., Lamm, R., van Steensel, B., and Becker, P.B. (2006). Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex. *Genes Dev.* 20, 858–870.
- Grskovic, M., Hentze, M.W., and Gebauer, F. (2003). A co-repressor assembly nucleated by Sex-lethal in the 3'UTR mediates translational control of *Drosophila* msl-2 mRNA. *EMBO J.* 22, 5571–5581.
- Hendrich, B., and Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.* 19, 269–277.
- Hilfiker, A., Hilfiker-Kleiner, D., Pannuti, A., and Lucchesi, J.C. (1997). mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila*. *EMBO J.* 16, 2054–2060.
- Hollmann, M., Simmerl, E., Schafer, U., and Schafer, M.A. (2002). The essential *Drosophila melanogaster* gene wds (will die slowly) codes for a WD-repeat protein with seven repeats. *Mol. Genet. Genomics* 268, 425–433.
- Huyen, Y., Zgheib, O., Ditullio, R.A., Jr., Gorgoulis, V.G., Zacharatos, P., Petty, T.J., Sheston, E.A., Mellert, H.S., Stavridi, E.S., and Halazonetis, T.D. (2004). Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* 432, 406–411.
- Kelley, R.L., Wang, J., Bell, L., and Kuroda, M.I. (1997). Sex lethal controls dosage compensation in *Drosophila* by a non-splicing mechanism. *Nature* 387, 195–199.
- Kind, J., Vaquerizas, J.M., Gebhardt, P., Gentzel, M., Luscombe, N.M., Bertone, P., and Akhtar, A. (2008). Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in *Drosophila*. *Cell* 133, 813–828.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705.
- Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410, 116–120.
- Legube, G., McWeeney, S.K., Lercher, M.J., and Akhtar, A. (2006). X-chromosome-wide profiling of MSL-1 distribution and dosage compensation in *Drosophila*. *Genes Dev.* 20, 871–883.
- Lemon, B., and Tjian, R. (2000). Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev.* 14, 2551–2569.
- Li, X., Wu, L., Corsa, C.A., Kunkel, S., and Dou, Y. (2009). Two mammalian MOF complexes regulate transcription activation by distinct mechanisms. *Mol. Cell* 36, 290–301.
- Lyko, F., Ramsahoye, B.H., and Jaenisch, R. (2000). DNA methylation in *Drosophila melanogaster*. *Nature* 408, 538–540.
- Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Mueller, J., et al. (2006). Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Mol. Cell* 21, 811–823.
- Morales, V., Straub, T., Neumann, M.F., Mengus, G., Akhtar, A., and Becker, P.B. (2004). Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. *EMBO J.* 23, 2258–2268.
- Nybakken, K., Vokes, S.A., Lin, T.Y., McMahon, A.P., and Perrimon, N. (2005). A genome-wide RNA interference screen in *Drosophila melanogaster* cells for new components of the Hh signaling pathway. *Nat. Genet.* 37, 1323–1332.
- Orphanides, G., and Reinberg, D. (2000). RNA polymerase II elongation through chromatin. *Nature* 407, 471–475.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17, 1030–1032.
- Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat. Genet.* 38, 700–705.



- Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858.
- Smith, E.R., Pannuti, A., Gu, W., Steumagel, A., Cook, R.G., Allis, C.D., and Lucchesi, J.C. (2000). The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Mol. Cell. Biol.* **20**, 312–318.
- Smith, E.R., Allis, C.D., and Lucchesi, J.C. (2001). Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in *Drosophila* males. *J. Biol. Chem.* **276**, 31483–31486.
- Taipale, M., Rea, S., Richter, K., Vilar, A., Lichter, P., Imhof, A., and Akhtar, A. (2005). hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol. Cell. Biol.* **25**, 6798–6810.
- Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41**, 105–178.
- Worby, C., Simonson-Leff, N., and Dixon, J.E. (2001). RNA interference of gene expression (RNAi) in cultured *Drosophila* cells. *Sci. STKE* **95**, pl1.
- Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H., and Allis, C.D. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859–872.
- Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* **15**, 2343–2360.
- Zhou, Y., Schmitz, K.M., Mayer, C., Yuan, X., Akhtar, A., and Grummt, I. (2009). Reversible acetylation of the chromatin remodelling complex NoRC is required for non-coding RNA-dependent silencing. *Nat. Cell Biol.* **11**, 1010–1016.



## **Appendix C**

**“The MOF chromobarrel domain controls genomewide H4K16  
acetylation and spreading of the MSL complex.”**

Thomas Conrad, Florence Cavalli, Erinc Hallaceli, Herbert Holz, Jop Kind,  
Ibrahim Ilik, Juan M. Vaquerizas, Nicholas M. Luscombe and Asifa Akhtar<sup>1</sup>

This manuscript was currently under review for publication in *Genes and Development* at the time of submission of this thesis.



# **The MOF chromobarrel domain controls genomewide H4K16 acetylation and spreading of the MSL complex.**

Thomas Conrad<sup>1,2</sup>, Florence Cavalli<sup>3</sup>, Erinc Hallacii<sup>1,2</sup>, Herbert Holz<sup>1</sup>, Jop Kind<sup>2,4</sup>, Ibrahim Ilik<sup>1,2</sup>, Juan M. Vaquerizas<sup>3</sup>, Nicholas M. Luscombe<sup>2,3</sup> and Asifa Akhtar<sup>1,\*</sup>

\*: Corresponding author

Email: [akhtar@immunbio.mpg.de](mailto:akhtar@immunbio.mpg.de)

Phone: (49) 07615108565

Fax: (49) 07615108566

Addresses:

1: Max-Planck-Institute of Immunobiology and Epigenetics, Stübeweg 51, 79108 Freiburg im Breisgau, Germany.

2: Genome Biology unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

3: EMBL European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK.

4: Present address: Netherlands Cancer Institute, Gene Regulation (B4), Plesmanlaan 121, 1066 CX Amsterdam.

**Running title:** MOF chromobarrel domain controls H4K16ac

**Keywords:** Dosage compensation, MSL, roX, transcription, acetylation, X chromosome.

## Abstract

A hallmark of dosage compensation in *Drosophila melanogaster* is MOF-mediated hyperacetylation of the single male X chromosome at lysine 16 of histone H4 (H4K16ac). MOF is recruited to X-linked target genes as part of the male specific lethal (MSL) complex, but has recently also been found at autosomal gene promoters. While chromodomains are known as recognition modules for methylated lysine residues, the MOF chromo-like “chromobarrel” domain interacts with nucleic acids. In this study, we reveal that the MOF chromobarrel domain is essential for H4K16 acetylation in the *Drosophila* genome. Furthermore, spreading of the MSL complex on the male X chromosome is abolished in chromobarrel domain mutants. The MOF chromobarrel domain therefore has an unprecedented function to potentiate MOF's enzymatic *in vivo*, making it the first example of a chromo-like domain directly controlling acetylation activity. Interestingly, while the chromobarrel domain is conserved in human MOF, we also find that the *Drosophila* specific N-terminal region of the MOF protein has evolved to perform sex specific functions. It modulates MOF's HAT activity and controls assembly of the MSL complex, thus regulating MOF function in dosage compensation. We propose that the MOF protein has been especially tailored to achieve tight regulation of its enzymatic activity to enable its dual role on X and autosomes.

## Introduction

Covalent modifications of histone tails modify chromatin structure to control transcription in response to environmental cues and developmental programs. Acetylation of lysine 16 on histone H4 (H4K16ac) has the potential to create or obscure binding platforms for chromatin modifying enzymes and transcriptional activators (Kim et al., 2010; Zippo et al., 2009). Furthermore, H4K16ac is the only histone mark that can directly impact on higher order chromatin structure by physically preventing formation of the 30nm chromatin fiber, thus creating an open, highly accessible chromatin environment (Bell et al., 2010a; Shogren-Knaak et al., 2006; Zippo et al., 2009). Indeed, H4K16ac is involved in a variety of chromatin related processes, such as replication timing and transcription (Akhtar and Becker, 2000; Bell et al., 2010b). Accordingly, the distribution of this histone mark needs to be tightly controlled throughout the genome, and loss of H4K16ac is often associated with human cancer (Fraga et al., 2005).

Global H4K16ac is also hallmark of the dosage compensated male X chromosome in *Drosophila*, where the resulting permissive chromatin structure facilitates spreading of the dosage compensation complex (DCC), also known as the Male Specific Lethal (MSL) complex, and contributes to the twofold transcriptional activation of X-linked genes (Straub and Becker, 2007). Hyperacetylation of the X chromosome is mediated by *Drosophila* MOF, a close homologue of hMOF, which resides in the MSL complex together with at least four other proteins, male specific lethal 1-3 (MSL1-3) and maleless (MLE), as well as two non-coding RNAs on the X (roX1/2) (reviewed in (Hallaceli and Akhtar, 2009; Ilik and Akhtar, 2009)). In addition to its role in dosage compensation, MOF has recently been found at hundreds of active promoters across the whole genome in male and female cell lines, where it is bound as part of the newly identified Non-Specific Lethal (NSL) complex (Kind et al., 2008; Mendjan et al., 2006; Raja et al., 2010). While several studies have found H4K16ac at the 5' end of genes in *Drosophila* (Bell et al., 2007; Kind et al., 2008; Schwaiger et al., 2009), it has remained unclear if MOF is acting as a histone acetyltransferase (HAT) at these sites, or if it has indeed any chromatin related function outside the context of the male X chromosome (Gelbart et al., 2009). Furthermore, it is not known how MOF targeting and activity are differentially regulated and distributed between the NSL complex and the MSL complex.

In addition to DNA binding domains, many components of the transcriptional machinery and other chromatin modifying factors contain specialized adapter modules that recognize specific histone modifications. Chromodomains are well known targeting modules that bind to methylated lysine residues. For instance, the chromodomain of HP1 has been shown to bind methylated lysine 9 of histone H3 (H3K9me3) (Lachner et al., 2001). Within the MSL complex there are two proteins that show similarities to chromodomains, MSL3 and MOF. The chromodomain of *Drosophila* MSL3 has been shown to interact with DNA and monomethylated H4K20 (Kim et al., 2010; Moore et al., 2010). Accordingly, mutations in the MSL3 chromodomain lead to reduced male viability, most likely due to defects in dosage compensation (Buscaino et al., 2006; Kim et al., 2010; Moore et al., 2010). A common feature of enzymes from the MYST family of HATs, such as *Drosophila* and human MOF, is the presence of a chromo-like domain, also known as the chromobarrel domain, adjacent to the enzymatic part of the protein (Sanjuan and Marin, 2001). However, in this case it lacks the aromatic cage necessary for methyl lysine binding (Nielsen et al., 2005). Instead, the MOF chromobarrel domain is required for MOF binding to roX RNAs *in vitro* and *in vivo* (Akhtar et al., 2000). In the absence of both roX RNAs or the RNA helicase MLE, MSLs including H4K16ac, are delocalized from the X chromosome and ectopically targeted to autosomal sites, the chromocenter and the 4<sup>th</sup> chromosome (Gu et al., 1998; Meller and Rattner, 2002). In line with these results, MOF is delocalized from the X chromosome upon treatment of permeabilized nuclei with RNase A (Akhtar et al., 2000). Remarkably however, ten years after the initial observation that chromobarrel domains are nucleic acid binding modules, their function and biological significance has remained elusive. Indeed, while nucleic acid binding of the chromobarrel domain has also been shown for yeast Esa1, no biological function has been described for chromobarrel domains of the MOF-type in general, despite their conservation from yeast to human (Sanjuan and Marin, 2001).

In this study, we show that *Drosophila* MOF is the major H4K16ac specific HAT across the male and female genome. Strikingly, disruption of the MOF chromobarrel domain leads to a genomewide loss of H4K16ac and consequently compromised MSL targeting to X-linked genes. Accordingly, we show that the chromobarrel domain serves to trigger H4K16ac after MOF binding to chromatin *in vivo* and *in vitro*. We thus reveal the biological role of the chromobarrel domain, which acts as an



accessory module to elicit the enzymatic capacity of its associated HAT enzyme. Furthermore, we also discovered that the N-terminal region of the MOF protein, which is absent in other organisms with different dosage compensation systems, controls assembly of the MSL complex on the male X chromosome and modulates MOF's HAT activity. Multiple levels of control therefore reflect the enhanced complexity of MOF functions in flies and the resulting need for increased context dependent regulation of MOF activity.

## Results

### MOF is required for genomewide H4K16ac and survival in both sexes

In this study, we set out to investigate the function of MOF in male and female flies. We were especially interested in determining the extent and biological significance of MOF activity inside and outside the context of the male X chromosome. In order to be able to manipulate MOF function in its *in vivo* context during downstream experiments, and to avoid the inherent limitations of cell line based approaches, we first wanted to obtain comprehensive genome-wide profiles from wildtype male and female flies. To this end, we first used antibodies against MOF and H4K16ac in chromatin immunoprecipitations, followed by ChIP-Seq from 3<sup>rd</sup> instar larva salivary glands (Table S1). The resulting profiles showed a very good overlap between sites of MOF binding and H4K16ac in both sexes (Figure S1A). On the male X chromosome, MOF binds to the entire gene body, with peaks of increased binding at the promoter and the 3'-end of genes. On male autosomes and generally in females, MOF binds only at gene promoters, as shown by composite MOF and H4K16ac profiles of average log2 fold-change values across all genes (Figure S1B). The pattern of H4K16 acetylation closely follows that of MOF binding. Interestingly, we observe a shift of the H4K16ac signal downstream of gene promoters, most likely reflecting acetylation of the first nucleosome downstream of the transcription start site (TSS). Strikingly, more than three quarters of all active genes were revealed as MOF targets on X and autosomes in both sexes (Table S2), signifying the role of MOF for genomewide gene regulation. Furthermore, there was a good correlation between MOF binding and H4K16ac, since most MOF bound genes were acetylated at the same time in both sexes (Figure S1D). However, on the male X chromosome, we also detected 735 genes that were acetylated despite lacking MOF, suggesting that transient the MSL complex interactions can spread H4K16ac in large domains across the male X chromosome, even in the absence of detectable MSL complex binding. Finally, we found an extremely high overlap of MOF target genes between males and females, especially on autosomes, suggesting that outside the context of the MSL complex MOF functions very similar in male and female flies (Figure S1C). These results confirmed our previous analysis but, due to the increased depth and sensitivity of the ChIP-sequencing method, also revealed that we had substantially underestimated the number of MOF targets (Figure S1E) (Kind et al., 2008).

We were next interested to test if MOF is active as a HAT also outside the context of the male X chromosome. For this purpose, we performed immunostainings of male and female 3<sup>rd</sup> instar larva polytene chromosomes from wildtype (wt) and *mof2* mutant flies that carry a premature stop codon and lack a functional MOF protein (Gu et al., 1998), using antibodies against MOF, MSL1 and H4K16ac. While the male X chromosome, marked by MSL1 staining, appeared enriched for H4K16ac in wt flies, we also detected widespread acetylation across autosomes as well as the chromocenter (Figure 1A). Likewise, female samples showed widespread H4K16ac, but no apparent enrichment on the X chromosomes, which also lacked MSL1 staining. Strikingly, H4K16ac was entirely lost from all chromosomes in male and female *mof2* flies, in the absence of MOF. In the males, this was accompanied by reduced staining of MSL1 on the X chromosome in a pattern most likely corresponding to the previously described high affinity sites (HAS), and subsequent delocalization of MSL1 staining to autosomal sites. Importantly, H4K16ac was restored on all chromosomes in both sexes upon expression of an HA-tagged MOF full length transgene (FL-MOF) in flies of the *mof2* background. We recapitulated the same global MOF dependency of H4K16ac by western blots using extracts prepared from 3<sup>rd</sup> instar larvae (Figure 1B). Female wt flies showed about half the amount of global H4K16ac compared to wt males, reflecting the presence of the hyperacetylated male X chromosome. In the absence of endogenous MOF in the *mof2* background, H4K16ac levels were drastically reduced in males and females, while no significant differences were observed upon probing with antibodies against H4. Importantly, H4K16ac could again be globally restored in both sexes by expression of FL-MOF. This data clearly demonstrates that MOF is responsible for genomewide H4K16ac in *Drosophila*.

Since MOF has been identified in a screen for male specific lethality, the obvious question arises, what the significance of MOF mediated H4K16ac is for female flies. To address this question we generated homozygous *mof2* females. When compared to control flies, the number of *mof2* females reaching adulthood was reduced by approximately two-fold, and these flies were mostly sterile (Figure 1C). More strikingly however, the average lifespan of *mof2* mutant female flies was drastically reduced to an average of 8 days as compared to 37 days in the control flies with the same set of balancer chromosomes (Figure 1D). Furthermore, this reduction in lifespan was rescued efficiently by expression of a MOF transgene. This result

strongly suggests that MOF mediated H4K16ac is essential for female survival, although it remains possible that additional MOF functions contribute to this phenotype.

### **The MOF chromobarrel domain and N-terminus are required for male viability**

Having established the general requirement of MOF for genome wide H4K16ac in male and female flies, we set out to investigate how the MOF chromobarrel domain contributes to this activity. Two point mutations in the MOF chromobarrel domain at Tyr416 (Y416) and Trp426 (W426) were previously shown to disrupt MOF's interaction with roX RNA (Akhtar et al., 2000), and the corresponding residues have subsequently been shown to be also required for nucleic acid binding in Esa1, a closely related HAT in yeast (Figure S2). Indeed, bacterial expressed wt chromobarrel domain directly interacted with a single stranded RNA probe in an electromobility gel shift assay (EMSA), and this interaction was lost upon mutation of Tyr416 or Trp426 (data not shown). To test if the chromobarrel domain carries essential functions *in vivo*, we assayed the capability of a series of MOF mutant proteins to rescue the male lethal phenotype associated with the loss of endogenous MOF. To this end, we introduced HA-tagged MOF transgenes by p-element mediated transformation into flies, including a control construct comprising full length MOF (FL-MOF); a deletion of the zinc finger region, which has been previously shown to be required for MOF binding to chromatin ( $\Delta$ Zn MOF ( $\Delta$ 565-587)); a series of deletions in the N-terminal half of the MOF protein ( $\Delta$ 1 ( $\Delta$  176-228),  $\Delta$  2 ( $\Delta$  241-357),  $\Delta$  3 ( $\Delta$  98-357) and  $\Delta$ N ( $\Delta$  1-349)); as well as two point mutations in the MOF chromobarrel domain (Y416D and W416G) (Figure S3). All transgenes were expressed to wt levels upon induction with armadillo-Gal4, as determined by western blot analysis (Figure S3 B-I). Full length HA-tagged MOF completely rescued male lethality (99.7% survival) in this assay when compared to heterozygous *mof2* females resulting from the same cross (Figure 2A), while males lacking the MOF zinc finger did not survive the 3<sup>rd</sup> instar larva stage. Male viability was also increasingly compromised by up to ~90% upon progressive deletions of the *Drosophila* specific N-terminal region of the MOF protein, suggesting that vital functions reside in this region. Strikingly however, disruption of the chromobarrel domain in Y416D and W426G MOF led to complete male lethality in this assay. This result clearly reveals the essential nature of the MOF chromobarrel domain that had been unprecedented.

## **The MOF N-terminus is required for the MSL complex assembly on the X chromosome**

Having confirmed the essential nature of the MOF chromobarrel domain and the importance of the N-terminal region for MOF function, we were next curious to assay for defects in MOF targeting to X-linked genes or autosomes, and in the MSL complex recruitment to the X chromosome. We performed immunostainings of 3<sup>rd</sup> instar larva polytene chromosomes from male flies that express MOF transgenes in the *mof2* background. Upon immunostaining with anti-HA antibodies, FL-MOF appeared in a wildtype pattern, showing pronounced enrichment on the X chromosome but also clear targeting to all autosomes (Figure 2B). At the same time, MSL1 staining remained restricted to the X, all together reflecting wildtype MSL complex targeting and function. This data provided an important control and confirmed the functionality of the full length HA-tagged MOF protein *in vivo*. As expected,  $\Delta$ Zn MOF was no longer detectable on polytene chromosomes, while at the same time immunostaining of whole mount salivary glands verified the nuclear localization of the  $\Delta$ Zn MOF protein, suggesting a general defect in chromatin binding (Figures 2C and S4) (Kadlec et al., 2011). Interestingly, upon disruption of the chromobarrel domain, Y416D and W426G MOF were still detected across all autosomes in immunostainings, however, MSL spreading on the X chromosome appeared reduced compared to FL-MOF (Figure 2D and see below). Strikingly however, in the absence of the MOF N-terminus, preferential targeting of MOF to the male X chromosome was entirely lost. Also MOF staining at autosomal bands appeared reduced (Figure 2E). Furthermore, we detected ectopic binding of MOF to the chromocenter. At the same time MSL1 staining was reminiscent of the patterns that have been observed in the absence of both roX RNAs (Meller and Rattner, 2002), with a nearly complete delocalization of MSLs from the X chromosome, accompanied by severe ectopic binding to autosomal sites and the chromocenter. Indeed, when we measured the levels of roX RNAs in the  $\Delta$ N mutant background we found a severe depletion of roX2 by more than 98%, suggesting that the N-terminus of MOF is required for proper incorporation of roXs into the MSL complex (Figure S5A). This phenotype was surprising and suggested that the N-terminus of the MOF protein in *Drosophila* species, which is not present in human MOF, is required for X chromosome and the MSL complex specific MOF functions. Consistent with MSL

binding at X chromosomal sites, we did not observe the same reduction in roX levels upon disruption of the chromobarrel domain (Figure S5A).

### **The chromobarrel domain is required for MSL spreading on the X chromosome**

It has been observed previously that the resolution provided by polytene chromosome stainings is not sufficient to detect defects in MSL spreading onto dosage compensated genes (Sural et al., 2008). It seemed therefore plausible that the lethality observed in chromobarrel domain mutants was caused by targeting defects on the individual gene level. Therefore, to get a more detailed insight into the MSL binding pattern upon disruption of the chromobarrel domain, we performed ChIP from male 3<sup>rd</sup> instar larva to monitor defects in MSL recruitment at higher resolution. Chromatin was immunoprecipitated, using antibodies against MSL1 and MSL3. The recovered DNA was measured by quantitative real-time PCR (qPCR). To monitor MSL recruitment to the X chromosome, we used the known high affinity sites (HAS) at the roX2 gene and an additional HAS at the cytological location 15A8, which had previously been identified in an MSL3 mutant background (Alekseyenko et al., 2008). We also included sites at the promoter, middle and 3'end of three X-linked genes, *Rpl22*, *Klp3a* and *Ucp4a*. The first one of these, *Rpl22*, previously showed some MSL binding in the absence of MSL3 and can thus be described as a medium affinity site, while the remaining two genes are low affinity MSL targets (Alekseyenko et al., 2008). As expected, MSL1 binding was retained on roX2 high affinity sites but was lost from the body of X-linked genes in the *mof2* background. These defects were restored back to the wt pattern upon expression of FL MOF (Figure 3A). Interestingly, the 15A8 HAS, which had been identified as a high affinity site in an MSL3 mutant background, was no longer bound by MSL1 in *mof2*, demonstrating the qualitative differences among high-affinity sites in varying genetic backgrounds (Kadlec et al., 2011).  $\Delta$ Zn MOF was unable to rescue the *mof2* phenotype (Figure S6A). Furthermore, MSL1 and MSL3 binding was lost from dosage compensated genes in the presence of  $\Delta$ N MOF, confirming immunostainings (Figure 3A). However, in contrast to the *mof2* background, where residual MSL1 and MSL2 containing complexes can bind to HAS, MSL1 binding was also lost from the roX2 HAS in  $\Delta$ N MOF, most likely as a result of compromised MSL complex assembly. Surprisingly, upon disruption of the chromobarrel domain in Y416D MOF, we found a dramatic reduction of MSL3 binding from X-linked target sites, particularly in the

transcribed region of genes (Figure 3B). At the same time, MSL1 binding was restricted to HAS and gene promoters, reminiscent of the pattern observed in the *mof2* background when MOF is absent. This result was striking and suggested a crucial role of the MOF chromobarrel domain for MSL targeting into the transcribed region of dosage compensated genes on the male X chromosome, consistent with the male lethality associated with disruption of this domain.

### **The chromobarrel domain is required for genomewide H4K16ac**

In addition to the consequences for MSL spreading on the X chromosome, we next wanted to more specifically address the role of the MOF chromobarrel domain and N-terminus for MOF function itself. We therefore performed immunostainings of polytene chromosomes using anti-H4K16ac antibodies. The pattern of H4K16ac in the presence of FL-MOF corresponded to the one observed in the wt, showing widespread acetylation on autosomes and enriched signal on the male X chromosome (Figure 2B). Acetylation was lost genomewide upon deletion of the zinc finger region in the MOF HAT domain, reflecting defective chromatin targeting of this mutant MOF protein (Figure 2C). To our great surprise however, we also saw a dramatic loss of the H4K16ac mark across all chromosomes in the presence of both chromobarrel domain mutants (Figure 2D). This was particularly striking since MOF binding to chromatin seemed to be much less affected. This data suggested a much more general role of the chromobarrel domain as we had anticipated, which seems to involve the activation of MOFs enzymatic capacity after its recruitment to chromatin, irrespective of the chromosomal context. In contrast, although autosomal chromatin targeting of  $\Delta N$  MOF appeared reduced in immunostainings, we detected widespread H4K16ac on all chromosomes in the presence of the  $\Delta N$  MOF protein, which contains a functional chromobarrel domain (Figure 2E).

Considering the dramatic reduction of H4K16ac upon disruption of the MOF chromobarrel domain, we wanted to assay for defects in chromatin targeting at higher resolution. To this end, we performed ChIP from Y416D MOF expressing male 3<sup>rd</sup> instar larva to monitor differences in MOF binding and H4K16ac. In addition to X-linked target sites we also assayed the promoter regions of the autosomal genes *cg6729*, *cg6884*, *cg31866*, *cg2708* and *cg7638* and at two non-targets, *bt* and *cg3937*. The pattern of FL-MOF binding in the *mof2* background was indistinguishable from the one observed for endogenous MOF in the wt, with clear

binding to autosomal promoters, to the promoter and transcribed region of X-linked genes and to HAS (Figure 4A). MOF binding was accompanied by H4K16ac at all of these loci in the wt and FL-MOF background. Upon removal of endogenous MOF in the *mof2* background alone, MOF binding and H4K16ac were both lost from all sites tested. Consistent with immunostainings,  $\Delta$ Zn MOF was unable to rescue this phenotype (Figure S6C). The binding pattern of  $\Delta$ N MOF suggested a general defect in chromatin targeting, showing a pronounced loss from X-linked target sites and a substantial reduction in promoter binding (Figure 4A). At the same time however, H4K16ac appeared much less affected on the same target sites (Figure 4B). At target promoters where  $\Delta$ N MOF binding was about twofold reduced, H4K16ac remained nearly at the levels observed in the presence of FL-MOF, suggesting a potentially enhanced acetylation activity of the truncated enzyme.

Disruption of the chromobarrel domain had diverse and contrasting effects on MOF targeting. The Y416D mutant MOF protein showed no difference in binding to gene promoters and the roX2 HAS (Figure 4C). However, binding to the 15A8 HAS and the transcribed region of *Rpl22* was twofold reduced, while targeting to the transcribed region of the low affinity genes *Klp3a* and *UCP4a* was completely lost. This data indicated a specific defect in chromatin targeting of Y416D MOF as part of the MSL complex, while the MSL complex independent binding to gene promoters appeared unaffected. We next analyzed H4K16ac across X-linked and autosomal target sites in the presence of Y416D MOF. Strikingly, H4K16ac was strongly reduced across all sites, including gene promoters and HAS, where the Y416D MOF protein was still readily detected (Figure 4D). This result suggests that the chromobarrel domain serves to trigger MOFs catalytic activity after initial recruitment of MOF to its chromatin targets.

MOF at gene promoters is part of the NSL complex (Raja et al., 2010). We therefore asked if reduced promoter binding of  $\Delta$ N MOF reflects compromised integration into the NSL complex. To address this question, we generated SL-2 cell lines stably expressing Flag-tagged FL MOF,  $\Delta$ N MOF and Y416D transgenes. After immunoprecipitation with anti-Flag antibodies followed by western blotting, membranes were probed against Flag, NSL1, NSL3 and MCRS2, as well as MSL1, MSL2 and MSL3. In this assay, FL MOF was stably interacting with all other complex partners (Figure S7). Also the binding of Y416D MOF to other MSLs appeared only slightly reduced, while interactions to NSL complex members remained unaffected.



Similarly, deletion of the MOF N-terminus did not have an apparent effect on the interactions with individual MSL proteins. Importantly however, interactions with all three members of the NSL complex were clearly diminished, consistent with reduced targeting of  $\Delta N$  MOF to gene promoters.

### **The chromobarrel domain triggers H4K16ac after nucleosome binding**

The observed loss of H4K16ac in chromobarrel domain mutants was striking and surprising. We therefore performed western blots from 3<sup>rd</sup> instar larva extracts to confirm the global reduction in H4K16ac by an independent method. When probing with HA antibodies, we found that all five MOF transgenes were stably expressed to similar levels (Figure 5A). Antibodies against tubulin and unmodified histone H4 were used as loading controls. MSL1 levels appeared clearly diminished in the presence of  $\Delta Zn$ ,  $\Delta N$ , Y416D and W426G MOF, reflecting compromised MSL complex assembly on X linked genes. MSL3 seemed less destabilized in the presence of Y416D MOF, but mirrored MSL1 levels in all other mutant backgrounds. FL-MOF efficiently rescued the *mof2*-mediated loss of H4K16ac to wt levels. No H4K16ac was detected upon deletion of the zinc finger in  $\Delta Zn$  MOF, reflecting compromised chromatin targeting of this mutant. Consistent with our previous analysis, substantial amounts of H4K16ac could be detected after deletion of the N-terminus in  $\Delta N$  MOF. Strikingly however, global H4K16ac levels were dramatically reduced in both chromobarrel domain mutants, confirming the general role of this domain for genome-wide H4K16ac. To control for the specificity of the assay we also probed against H4K5ac as well as H4K8ac and H4K12ac. These histone marks are not mediated by MOF in *Drosophila*, and remained unchanged in each of the MOF mutants analyzed (Figure 5A and data not shown).

We next wanted to recapitulate the requirement of the chromobarrel domain for MOF HAT activity *in vitro*. To this end we performed *in vitro* HAT assays on purified endogenous nucleosomes, using baculovirus expressed FL MOF, Y416D, W426G and  $\Delta N$  MOF that were copurified with MSL1 and MSL3 to yield enzymatically active trimeric complexes (Morales et al., 2004). Consistent with our *in vivo* observations, H4 directed acetylation activity was approximately three to five-fold reduced in trimeric complexes containing Y416D and W426G MOF, as compared to FL MOF (Figure 5B and data not shown). We also observed that  $\Delta N$  MOF containing trimeric complexes showed an about five-fold increase in acetylation activity compared to FL

MOF, while monomeric  $\Delta N$  MOF showed the same low levels of residual activity (Figure 5C). This result suggests that an autoregulatory function in the N-terminal domain of MOF constrains its enzymatic activity, and explains why levels of H4K16ac remain high *in vivo* although targeting of  $\Delta N$  MOF to chromatin is impaired.

The reduced acetylation activity of chromobarrel domain mutant MOF was specific to nucleosomal substrates, since the same trimeric complexes showed comparable activities towards free histones, which also confirmed functionality of the MOF HAT domain itself (Figure 5D). The activity of  $\Delta N$  MOF containing trimeric complexes towards free histones was slightly enhanced, although the increase was not as pronounced as on a nucleosomal substrate. We then wanted to ask at which step of initial substrate binding and subsequent acetylation the chromobarrel domain functions. To address this question we performed affinity purifications of trimeric complexes using biotinylated DNA or mononucleosomes coupled to streptavidin beads. Nucleosome binding of trimeric complexes containing the  $\Delta N$  MOF protein was slightly weakened as compared to FL MOF (Figure 5E). This result further highlights the enhanced enzymatic activity of  $\Delta N$  MOF on the same substrate. Strikingly, trimeric complexes containing Y416D and W426G MOF proteins bound as efficiently to the nucleosomal substrate as FL MOF trimeric complexes. Initial binding to the nucleosome is thus independent of a functional chromobarrel domain, which is then required to activate the subsequent acetylation of the H4 tail, the extent of which is controlled by the MOF N-terminus.

## Discussion

In this study we reveal a striking and unexpected role of the MOF chromobarrel domain for the control of genomewide H4K16ac and dosage compensation in *Drosophila*. We show that upon MOF recruitment to its target sites H4K16ac can only be triggered in the presence of a functional chromobarrel domain *in vivo* and *in vitro*. To our knowledge, this is the first example in which a chromodomain directly controls the activity of its associated HAT enzyme. At the same time the *Drosophila* specific N-terminus modulates overall acetylation levels and controls MSL complex assembly, thereby coordinating the dual role of MOF in flies.

### MOF is the major H4K16 specific HAT in *Drosophila*

Since the recent discovery that MOF resides at autosomal gene promoters as part of the NSL complex, the full extent of MOF function at these sites has remained elusive (Gelbart et al., 2009; Kind et al., 2008; Raja et al., 2010). Disruption of other NSL complex members leads to lethality in males and females (Mendjan et al., 2006). This suggested that MOF is not strictly necessary for NSL complex function, since adult female flies can be recovered in the absence of MOF (Hilfiker et al., 1997). However, although MOF was identified in a screen for male specific lethality, this study did not address defects in female viability. We now show that MOF is indeed an essential gene in both sexes, since the number of females reaching adulthood, and especially female lifetime, are drastically reduced in the absence of MOF. Furthermore, our data demonstrates that, in addition to its role in X chromosome dosage compensation, MOF is targeted to the vast majority of all active gene promoters in male and female flies. Importantly, we also show that in the absence of MOF, H4K16ac is lost from all MOF target sites in the male and female genome. In fact, bulk H4K16ac is lost genome wide in 3<sup>rd</sup> instar larva upon disruption of MOF in both sexes, suggesting that MOF is the major H4K16ac specific HAT in *Drosophila*. However, since a recent study showed that ATAC2 is contributing to bulk H4K16ac during embryonic development (Suganuma et al., 2008), the possibility remains that additional enzymes might have the capacity to mediate H4K16ac in certain developmental stages or tissues. Interestingly, bulk H4K16ac is also strongly diminished in human cells upon RNAi and mutants of the close homologue hMOF (Gupta et al., 2008; Smith et al., 2005; Taipale et al., 2005; Thomas et al., 2008), and it has subsequently been shown that hMOF is targeted to thousands of gene promoters across the

human genome (Wang et al., 2009). It is noteworthy that dosage compensation in mammals does not involve MSL proteins, and accordingly hMOF does not appear enriched on the inactive X chromosome (Wang et al., 2009). The most closely related MYST domain in yeast resides in *S.cerevisiae* SAS2p (Sanjuan and Marin, 2001), which hyperacetylates subtelomeric regions to prevent Sir3p binding and spreading of telomeric heterochromatin (Kimura et al., 2002; Suka et al., 2002). However, a better resemblance of the MOF domain architecture is found in the yeast homologue Esa1p, which like MOF contains a chromobarrel domain preceding the MYST domain. Interestingly, Esa1p is generally recruited to the promoters of active protein coding genes where it mediates acetylation of histone H4 as part of the NuA4 complex, and this complex also contains a yeast homologue of *Drosophila* MSL3, Eaf3 (Robert et al., 2004). It is therefore highly likely that the pattern of MOF binding that we observe at promoters across male and female autosomes and on the female X chromosome in *Drosophila*, is indeed reflecting the most ancient mode of MOF function.

### **The MOF chromobarrel domain controls H4K16ac genomewide**

An earlier study claimed that the MOF chromobarrel domain had a minor role to play in dosage compensation (Morales et al., 2004). However, this study only tested the capacity of MOF to target to the X chromosomal territory by simply overexpressing MOF variants in a male cell line, leaving the possibility that potential defects in MOF function have been masked by the presence of the endogenous protein. Indeed, demonstrating the advantages of the *in vivo* system, we were able to reveal the crucial role that the MOF chromobarrel domain plays for all aspects of MOF function. Upon disruption of its nucleic acid binding properties, MSL spreading to X-linked genes is compromised leading to a defect in dosage compensation. This is reminiscent of the phenotype observed upon deletion of the chromobarrel domain of MSL3 (Sural et al., 2008). However, the MSL3 chromobarrel domain is thought to contribute to MSL targeting via its binding to the H3 tail trimethylated at K36 (Larschan et al., 2007). Also an interaction with the H4 tail monomethylated at K20 has been proposed (Kim et al., 2010; Moore et al., 2010). In contrast, the MOF chromobarrel domain lacks the aromatic cage required for binding to methylated lysine residues. We therefore believe that reduced MSL spreading in MOF chromobarrel domain mutants is not the result of a direct chromatin binding defect of

the MOF protein, consistent with the fact that binding of MOF to gene promoters *in vivo* and to nucleosomes *in vitro* is unaffected upon disruption of the chromobarrel domain. It has been shown previously that the chromobarrel domain mediates MOF binding to roX RNAs, and it is possible that this interaction contributes to spreading of the MSL complex along dosage compensated genes. However, the most dramatic and unexpected consequence upon disruption of the chromobarrel domain was the dramatic loss of genomewide H4K16ac. It is clear from previous work that this reduction of H4K16ac alone is sufficient to disrupt MSL spreading on the X chromosome. Indeed, upon direct mutation of the catalytic site in the MOF enzyme, MSL binding is restricted to high affinity sites in a pattern similar to the one observed in chromobarrel domain mutants (Gu et al., 1998).

The structure of the *Drosophila* MOF chromobarrel domain (aa 367-454) has been determined by NMR (Nielsen et al., 2005) (Figure S2). While Tyr416 (Y416) is one of the putative aromatic cage residues and is partially buried in the core of the structure, Trp426 (W426) is solvent exposed on a  $\beta$ -sheet formed by strands  $\alpha 2$ ,  $\alpha 3$  and  $\alpha 4$ . As mentioned above, a similar domain is present also in the yeast Esa1p. It has been shown there that in presence of a short N-terminal extension, a minor conformational change occurs in the core of the Esa1p domain, triggering its RNA/DNA binding activity (Shimojo et al., 2008). Interestingly, a similar structure of the chromobarrel domain of MSL3 has recently been determined in complex with dsDNA and the N-terminal tail of histone H4 monomethylated at H4K20 (Kim et al., 2010). Both MOF and Esa1 possess equivalent highly conserved surfaces that could be involved in the interaction with nucleic acids (Figure S2D-F). Accordingly, extensive mutagenesis studies in Esa1p identified residues involved in nucleic acid binding, including residues corresponding to MOF Tyr413, Tyr416, Asn420, Arg422 and Trp426. Mutations of these residues in Esa1p were lethal or produced severe growth effects (Shimojo et al., 2008). In the MSL3 structure, Trp66 corresponding to MOF chromobarrel domain Trp426 is directly involved in the interaction with DNA (Figure S2E). Although the aromatic residues that are critical for interaction of the MSL3 chromobarrel domain with the methylated lysine side chain of the H4 peptide are not well conserved, it remains unclear whether MOF and Esa1 could bind unmodified histone tail residues. However, one could potentially envisage that a mode of action similar to MSL3 may also exist for the Esa1p and MOF chromobarrel domains, and that interaction with nucleic acid and the unmodified H4 tail direct the tail for

acetylation by the MYST domain. It has been shown previously that the MOF chromobarrel domain binds to nucleic acids with a preference for RNA over DNA, and interacts with roX RNAs *in vivo* (Akhtar et al., 2000). We now show that the nucleic acid binding properties of the MOF chromobarrel domain are necessary to trigger H4K16ac on a nucleosomal substrate *in vitro*, where the local DNA concentration is high due to tethering of the complex to the nucleosome via other domains. Interestingly, deletion of a conserved stem loop structure in roX RNAs leads to a specific loss of hyperacetylation from the male X chromosome *in vivo*, while MOF is still targeted to the X chromosome in this background (Park et al., 2007; Park et al., 2008). It is therefore tempting to speculate that upon ordered assembly of roX RNAs and MSL proteins *in vivo*, roX RNAs could engage in similar interactions with the chromobarrel domain and H4 tail to promote high acetylation levels throughout the body of X-linked genes. Since the chromobarrel domain is also conserved in human MOF it is likely that MOF's acetylation activity is controlled in a similar manner in humans.

### **MOF controls dosage compensation via its N-terminal domain**

Another striking observation during this study was that the N-terminal part of the MOF protein is required for the MSL complex assembly, and in addition modulates the acetylation activity of MOF. The presence of a large MOF N-terminal region is a prominent feature of *Drosophila* species, and absent in all other MOF homologues from yeast to human (Figure S5B). Addition of this domain correlates with the evolution of the *Drosophila* dosage compensation system, involving the targeting of MSLs and H4K16ac to the male X chromosome (Bone and Kuroda, 1996). It seems thus plausible that the N-terminal domain has added functions that utilize the evolutionarily ancient transcriptional regulator MOF for the novel task of dosage compensation. Strikingly, upon deletion of the MOF N-terminus, MSL targeting to the X chromosome was completely abolished. This also included high affinity sites (HAS), which are otherwise resistant to MOF depletion (Gu et al., 1998). Deletion of the MOF N-terminus therefore has a dominant negative effect on MSL complex formation. Aberrant MSL targeting was reminiscent of the phenotypes observed in roX1 and roX2 double mutants (Meller and Rattner, 2002), and indeed roX levels were reduced by 98% in the absence of the MOF N-terminus. This result suggests that, although interaction with MSL1 and MSL3 is mediated by the zinc finger region

of MOF (Morales et al., 2004), the N-terminus is required to assemble the core MSL subunits together with roX RNAs into a functional MSL complex. At the same time, the N-terminus is also required for proper integration of MOF into the NSL complex at gene promoters. Importantly, despite these defects, N-terminally truncated MOF was still active as a HAT and able to target to chromatin (Figures 2, 4).

MSL protein domains required for the MSL complex assembly and targeting, as well as their cognate DNA binding sequences have been shown to undergo rapid adaptive coevolution in *Drosophila melanogaster* (Bachtrog, 2008; Rodriguez et al., 2007). Also roX RNA sequences are highly divergent throughout *Drosophila* species (Park et al., 2007). Likewise, the N-terminal domain of MOF shows huge variation in size and amino acid sequence between *Drosophila* species suggesting ongoing selective pressure by other MSLs or roX RNA. Another striking feature of the N-terminal domain is its intrinsic disorder, according to secondary structure prediction. It is a recently emerging concept that protein function is not necessarily linked to fixed secondary and tertiary structures (Chouard, 2011). Functional disordered regions have been particularly suggested for domains in hub proteins that control a variety of biological processes and mediate interactions to multiple interaction partners (Oldfield et al., 2008). The most prominent example of this type of regulation is p53, which interacts with hundreds of binding partners via intrinsically unstructured domains. The N-terminal domain of MOF has relatively few hydrophobic amino acids, suggesting solubility in solution. We therefore propose that the unstructured N-terminal domain integrates the multiple functions of MOF in dosage compensation and for genomewide H4K16ac (Figure 6).

A further intriguing finding was that the N-terminus constrains MOF HAT activity *in vivo* and *in vitro*. Again the necessity for this additional level of regulation arises from the dual function that MOF has adopted in *Drosophila*. We show that H4K16ac displays much higher baseline levels on the male X chromosome as in any other chromosomal context. Furthermore, MOF mediated H4K16ac extends beyond regions of MSL binding on the male X chromosome. We do not observe the same phenomenon on male autosomes or in females, where much lower levels of H4K16ac are restricted to sites of MOF binding at gene promoters (Figure S1). The activity level of MOF thus differs to a large degree depending on the chromosomal context. It is known that MOF requires interaction with MSL1 and MSL3 to be active as a HAT *in vitro*. However, our data suggests that an additional constrain is imposed

on MOF's enzymatic activity by an autoinhibitory function residing in its N-terminus. Interestingly, our data shows that in the  $\Delta N$  MOF background, H4K16ac spreads from sites of MOF binding into neighboring regions even in the absence of other MSLs. We therefore propose that in the wildtype situation, the N-terminus may control or restrict the H4K16ac spreading around sites of MOF binding, and this constrain is only released in the presence of a fully assembled MSL complex on the male X chromosome for more extensive acetylation.

In previous experimental setups, loss of MOF mediated H4K16ac has been accompanied by simultaneous loss of MSLs from the transcribed regions of X-linked genes (Gelbart et al., 2009; Gu et al., 1998; Hilfiker et al., 1997; Kind et al., 2008). Therefore the possibility remained that H4K16ac could merely serve to allow MSL spreading along the X chromosome, and upregulation of transcription would be achieved subsequently by additional activities residing in other MSL proteins. Although MSLs were entirely lost from the body of low affinity target genes in the absence of the MOF N-terminus, the reduction in H4K16ac was less pronounced, most likely reflecting the presence of hyperactive MOF at nearby promoters. Intriguingly, about 11% of males escaped lethality in this background. This suggests that in these individuals the presence of H4K16ac on X chromosomal genes has been sufficient to ensure dosage compensation despite the lack of MSL binding. This result supports the view that H4K16ac is sufficient to upregulate transcription of X-linked genes, and that the main function of the MSL complex is to direct H4K16ac to the male X chromosome.

## Summary

In this study we have revealed the function of the MOF chromobarrel domain, which is to elicit the activity of its associated HAT enzyme on the chromatin target. The huge degree of sequence conservation of the MOF protein between flies and mammals suggests that a similar mode of operation might be present in mammalian MOF. Our findings thus have important implications for the study of this enzyme in humans, where it is implicated in a wide range of processes like transcription, DNA repair and cancer (Rea et al., 2007; Wang et al., 2009; Zippo et al., 2009). In addition, it will be exciting to test if chromodomains in other enzymatic complexes can directly control their associated enzyme activities in a similar way. In contrast to the chromobarrel domain, the N-terminal part of the MOF protein is specific to *Drosophila*



species. Accordingly, our work also highlights how a MYST HAT has been adopted through evolution to carry out distinct functions on X chromosomal and autosomal genes. We propose that the enzymatic activity of other MYST family HATs may be under similar regulation by associated domains for context specific function.

## Materials and Methods

### ChIP from salivary glands

100 Inverted 3rd instar larva were fixed in fixing solution (50 mM HEPES pH 7.6, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 3.6% formaldehyde) for 20 min at room temperature on a rotating wheel. Formaldehyde was quenched with stop solution (PBS, 0.01% Triton X-100 0.125 M glycine) for 5min. Fixed larva were washed for 5 x 2 min with buffer A (10 mM HEPES pH 7.6, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100) and 5 x 5 min with buffer B (200 mM NaCl, 10 mM HEPES pH 7.6, 1 mM EDTA, 0.5 mM EGTA, 0.01% Triton-X 100). Salivary glands were dissected into 500  $\mu$ L of RIPA buffer (25 mM HEPES pH 7.6, 150mM NaCl, 1 mM EDTA, 1% Triton-X 100, 0.1% SDS, 0.1% DOC, protease inhibitors) and sonicated 16 x 30 sec using a Branson 250 sonicator at 40 pulse, intensity 5, generating 200 bp fragments on average. The chromatin was cleared by 10 min of high speed centrifugation at 4°C in a table top centrifuge. The chromatin was then aliquoted in 8 samples and snap frozen. For immunoprecipitation, one aliquot was filled up with RIPA buffer to 600 $\mu$ L and incubated with 2 $\mu$ L of rabbit anti-MOF, 2  $\mu$ L rabbit anti-H4 (abcam ab10158), or 5 $\mu$ L rabbit anti-H4K16ac (Santa Cruz sc-8662-R) antibodies O/N at 4°C. 50  $\mu$ L of this material was then taken as Input control and 500  $\mu$ L used for immunoprecipitation. For ChIP-Seq, the complete material from 4 x 100 salivary glands was used for Immunoprecipitation, using 3 $\mu$ L of rabbit anti-MOF, 3  $\mu$ L rabbit anti-H4, 5 $\mu$ L rabbit anti-H4K16ac, or 6 $\mu$ L of rabbit anti-Rpb3 (gift from John Lis and Karen Adelman). Immunocomplexes then isolated by adding protein A -Sepharose (Roche) for 3 hours, followed by six washing steps: 4x RIPA buffer, 1x DOC buffer (10 mM Tris at pH 8, 0.25 M LiCl, 0.5% NP-40, 0.5% DOC, 1 mM EDTA), and 1x TE at pH 8. The beads were resuspended in 90  $\mu$ L TE and, together with the Input control, the crosslink was reversed at 65°C over night. After 30 min incubation at 37°C with RNaseA (0.2 mg/ml), followed by 2 hr Proteinase K digestion (0.05 mg/ml) at 50°C, DNA was purified using Minelute columns (Qiagen). ChIP DNA samples were resuspended in 500  $\mu$ L nuclease free water. We used 10  $\mu$ L ChIP material for each qPCR reaction. For ChIP-Seq, the material from four independent IPs was pooled for each experiment (two replicates were generated for all Rpb3 data sets).

### **ChIP from whole larva**

For whole larva ChIP, for each replicate, 100 larva were crushed to a powder in liquid nitrogen, dounced 30x in 20 ml NE buffer (15mM Hepes pH 7.6, 10mM KCl, 5mM MgCL<sub>2</sub>, 0.1mM EDTA, 0.5mM EGTA, 350mM sucrose, 0.1% Tween, 1mM DTT, protease inhibitors) and fixed by adding 1ml 36% formaldehyde to 1.8% final concentration for 20 min at RT. After quenching for 5 min at RT with 125 mM glycine, nuclei were collected by 10 min centrifugation at 2500g, washed 3x 5min in RIPA (25 mM HEPES pH 7.6, 150mM NaCl, 1 mM EDTA, 1% Triton-X 100, 0.1% SDS, 0.1% DOC, protease inhibitors) sonicated 5x 30 sec in a Branson 250 sonicator at 30 pulse, intensity 3, followed by sonication in a Covaris SL-2 sonicator for 6 min using the preset 200bp program. After clearing by high-speed centrifugation, the chromatin was split in 8 aliquots and snap frozen for subsequent immunoprecipitation. For immunoprecipitation, the volume was made up to 600 µl with RIPA. 2 µl of rabbit MSL1 or MOF, 3µl of rat MSL3, 5 µl of anti-H4K16ac (Santa Cruz sc-8662-R) 2 µl anti-H4 (abcam ab10158), 2 µl anti-H3 (ab1791) antibodies were added and incubated over night at 4°C, respectively. Immunocomplexes were isolated by adding protein A/G-Sepharose (Roche) for 3 h, followed by six washing steps: 4x RIPA buffer, 1x DOC buffer (10 mM Tris at pH 8, 0.25 M LiCl, 0.5% NP-40, 0.5% DOC, 1 mM EDTA), and 1x TE at pH 8. The beads were resuspended in 90 µl TE and, together with the Input control, the crosslink was reversed at 65°C O/N. After 30 min incubation at 37°C with RNaseA (0.2 mg/ml), followed by 2 hr Proteinase K digestion (0.05 mg/ml) at 50°C, DNA was purified using Minelute columns (Qiagen). ChIP DNA samples were resuspended in 500 µl nuclease free water. We used 10 µl ChIP material for each qPCR reaction.

### **RNA isolation and quantitative PCR analysis**

For qRT-PCR, RNA and corresponding genomic DNA from SL-2 cells or from salivary glands were isolated simultaneously using the AllPrep DNA/RNA mini kit (QIAGEN), the RNA column was DNaseI treated, and 300 ng of total RNA was used in an RT reaction. qPCR was performed on an Applied Biosystems (AB) Cyclor7500 with SYBR detection, and the amplification curves were analyzed with the corresponding AB software. Each qRT-PCR was repeated at least three times, values were normalized to corresponding genomic DNA values, and the standard deviation within each experiment was calculated. qPCR analysis of ChIP samples

was performed using the SYBR Green PCR master mix (Applied Biosystems), 100 ng of each forward and reverse primer, and 10 µl immunoprecipitated DNA, in an ABI7500 real-time PCR thermocycler (Applied Biosystems, Inc.). Recovery was determined as the amount of immunoprecipitated DNA relative to input DNA. Error bars represent standard deviation (StDev) of four independent experiments.

### **Fly stocks and crosses**

All stocks were maintained on standard medium at 25°C. To assay for female viability we conducted the cross *y/mof<sup>2</sup>;sb/P{w+ UAS-HA-MOF}c* x *mof<sup>2</sup>/mof<sup>2</sup>;P{w+ UAS-HA-MOF}/tm3* and the control cross *y/w-;sb/+* x *w-/w-;+/tm3* and compared the proportion of *mof<sup>2</sup>/mof<sup>2</sup>;sb/t3* or *w-/w-;sb/t3* in the respective offspring. For the complementation test, female flies of the genotype *mof<sup>2</sup>/fm7;P{w+ UAS-HA-MOF}* were crossed to *y/fm7;P{armadillo-GAL4}* males to induce transgene expression, and the ratio of male *Y/mof<sup>2</sup>* to female *mof<sup>2</sup>/fm7* offspring was scored as relative male survival.

For the in vivo characterization of MOF variants the following fly strains were generated:

1. *w; P{ w+ UAS-HA-MOF}/TM3*
2. *w; P{ w+ UAS-HA-□ 565-587-MOF}/TM6*
3. *w; P{ w+ UAS-HA-□ 565-587-MOF}/TM3*
4. *w; P{ w+ UAS-HA-□ 1-349-MOF}/Cyo*
5. *w; P{ w+ UAS-HA-□ 1-349-MOF}/TM3*
6. *w; P{ w+ UAS-HA-□ 1-349-MOF}/TM3*
7. *w; P{ w+ UAS-HA-Y416D-MOF}/Cyo*
8. *w; P{ w+ UAS-HA-Y416D-MOF}/Cyo*
9. *w; P{ w+ UAS-HA-Y416D-MOF}/TM6*
10. *w; P{ w+ UAS-HA-W426G-MOF}/TM3*
11. *w; P{ w+ UAS-HA-W426G-MOF}/Cyo*
12. *w; P{ w+ UAS-HA-□ 176-228-MOF}/*
13. *w; P{ w+ UAS-HA-□ 176-228-MOF}/*
14. *w; P{ w+ UAS-HA-□ 241-357-MOF}/*
15. *w; P{ w+ UAS-HA-□ 241-357-MOF}/*
16. *w; P{ w+ UAS-HA-□ 98-357-MOF}/Cyo*
17. *w; P{ w+ UAS-HA-□ 98-357-MOF}/Cyo*

### **Immunostaining of polytene chromosomes and confocal microscopy**

Two to three pairs of salivary glands were dissected in PBS for each squash. The glands were transferred into a drop of 35  $\mu$ l Fix solution (1.8% PFA, 45% Acetic acid, in H<sub>2</sub>O) on a glass slide and incubated for 10 min. A cover slip was put on top of the sample and tapped with a pencil to disrupt cells and nuclei, then heavily pressed with the thumb to spread the chromosomes. The slide was snap frozen in liquid nitrogen, the cover slip removed with a razor blade and the slide transferred into PBS. The slide was then incubated in blocking solution (3% BSA, 0.2% NP40, 0.2% Tween20, 10% dry milk, in PBS) for 1 h. After transfer into a wet box, 40  $\mu$ l blocking solution with primary antibody were added onto the sample and incubated for 1 h at RT or over night at 4°C. MOF and MSL1 antibodies were used at 1:500, anti-K16ac Santa Cruz at 1:125 dilutions. After two times rinsing with PBS, the slide was washed for 15 min at RT with washing buffer (500mM NaCl, 0.2% NP40, 0.2% Tween20, in PBS). The slide was then incubated in the dark with 40  $\mu$ l blocking solution containing a 1:250 dilution of fluorescently labeled secondary antibody. The slide was rinsed two times with PBS and washed for 15 min at RT with washing buffer. After another rinse with PBS, 15  $\mu$ l of Fluoromount were added onto the sample and the sample was covered with a cover slip. Images were captured with an AxioCamHR CCD camera on a Leica SP5 (Leica Microsystems) using an Apochromat NA 1.32 oil immersion objective. Images were arranged with Adobe Illustrator.

### **Generation of protein extracts and western blotting.**

To determine protein levels in adult fly heads, flies were first anesthetized under CO<sub>2</sub> and then sorted by sex. 50 flies were selected, transferred into an eppendorf tube and snap frozen in liquid nitrogen. Upon vigorous shaking of the tube, heads detach and can be collected into 100  $\mu$ l of 2 x Laemmli buffer. Heads were sonicated for 10 sec in a Branson 250 sonicator at 30 pulse, intensity 3. After 10 min centrifugation at full speed in a table top centrifuge, the supernatant was collected and boiled for 5 min at 95°C. 10  $\mu$ l of the sample were then used for PAGE. To determine protein levels in 3<sup>rd</sup> instar larva, 10 larvae were inverted in PBS, the gut, fat body and salivary glands were removed and the sample transferred into 100  $\mu$ l 2 x Laemmli buffer. After sonication for 10 sec in a Branson 250 sonicator at 30 pulse, intensity 3, the sample was boiled for 5 min at 95°C. The sonication was repeated and the

sample subsequently cleared by 10 min high speed centrifugation. 10 µl of the sample were used for PAGE. After wet transfer onto a PVDF membrane (Millipore) and 1 h incubation in blocking solution (5% milk, 0.3% Tween-20, PBS), membranes were probed for 1 h at RT with blocking solution containing 1:3.000 rabbit anti-MOF; 1:3.000 rabbit anti-MSL1; 1:1.000 rat anti-MSL3; 1:1.000 mouse anti-HA; 1:10.000 rabbit anti-H3; 1:1.000 rabbit anti H4; 1:1.000 rabbit anti-H4K16ac (Santa Cruz sc-8662-R); or rabbit anti-H4K5ac antibodies, respectively. Membranes were washed 3 x 5 min in PBS-T (0.3% Tween-20 in PBS) and incubated for 45 min with blocking solution containing 1:15.000 anti-mouse or anti-rabbit antibodies coupled to horseradish peroxidase. After another 3 x 5 min wash with PBS-T, membranes were incubated for 5 min with Lumi-Light western blotting substrate (Roche). The luminescence signal was captured on Kodak Biomax MR films.

### **Expression and purification of *Drosophila* recombinant proteins with the baculovirus system**

The MOF variant proteins expressed in the baculovirus system were cloned as full length constructs in HA-pFastBac vectors. Recombinant baculoviruses were generated as described in the manual “Bac-to-Bac Baculovirus Expression System” (Invitrogen). Recombinant viruses for expression of MSL1 and FLAG-MSL3 were readily available in the lab. Recombinant viruses were used to infect insect cells (SF21) and these were harvested 2 days after infection by dissolving the cell pellets in HEMG-200 Buffer (25mM Hepes pH 7,6 ; 1mM EDTA ; 12,5 mM MgCl<sub>2</sub> ; 10 % Glycerol ; 200 mM KCl, 0,5% Triton X-100 reduced ; Complete Protease Inhibitor(Roche) ; 0,2 µM PMSF). For purification of the recombinant proteins, whole-cell-extracts were incubated for 2 h with HA-Agarose or Flag-Agarose-Beads (Sigma), followed by washes with HEMG-500 and HEMG-200. HA-and Flag-peptides (Sigma) were used in a concentration of 400 ng/µl to elute the recombinant proteins. All recombinant proteins were stored in HEMG 200 Buffer.

### **HAT-assay on nucleosomal templates**

In a 20 µl reaction containing 50 mM Tris pH 8,0, 0,1 mM EDTA pH 8, 0, 1 mM DTT, 1 mM PMSF, 5% Glycerol, 10 mM Na-butyrate and 0,02 µCi of acetyl coenzyme A (acetyl-1-14C, 60 mCi/mmol (Perkin Elmer)) we incubated 5.3 nM (10 ng) of the respective recombinant MOF-protein (or trimeric complexes) with 1,5 µg of

endogenous mono, -di- nucleosomes obtained from MCF-7 cells (human breast cancer cell line). For preparation of MCF-7 nucleosomes: See Epigenome Network of Excellence (NoE) protocols (<http://www.epigenome-noe.net/researchtools/protocol.php?protid=22>). After a 60 min incubation at 26°C the whole reaction volume was applied on a precast 12% BIS/Tris Novex Gel (Invitrogen) and run in 1x MES Buffer at 130V for 45 min and 155 V for another 45 min. Proteins were visualized by Coomassie R 250 Blue staining, destained, and the gel was dried on a 3 MM paper using a vacuum dryer (Biorad) at 80°C for 2 h. The dried gel was subsequently exposed on an imaging plate BAS-IP MS 2025 for 5 days. Acetylation signals were obtained by scanning the IP-MS plate on a FLA5000 scanner (Fujifilm).

### **Filter binding HAT-assay**

In a 30 µl reaction containing 50 mM Tris pH 8,0, 0,1 mM EDTA pH 8, 0, 1 mM DTT, 1 mM PMSF, 5% Glycerol, 10 mM Na-Butyrate and 0,02 µCi of acetyl coenzyme A (acetyl-1-14C, 60 mCi/mmol (Perkin Elmer)) we incubated 21.3 nM (40 ng) of the respective recombinant MOF-protein (or trimeric complexes) with 1,5 µg of recombinant (Xenopus) histone octamer. After a 60 min incubation at 26°C the whole reaction volume was applied on a P81 filter paper (Whatman, 1,5 x 1,5 cm), air dried and washed 3x10 min at RT in a large volume of 50 mM Na-carbonate pH 9,2. After rinsing the filters in acetone, they were air dried and counted in 5 ml of scintillation liquid (Rotiszint eco plus, Roth).

### **Immunoprecipitation of MOF constructs from SL-2 stable cell lines**

All MOF constructs used for SL-2 stable cell line generation carry N-terminal 3xFlag 6His tags and are under the control of an MtnA promoter. The vector also contains a Neomycin cassette, allowing omitting of the co-transfection step. 1 million cells were transfected with 0.5 µg of DNA with Qiagen Effectene reagent. The cells were selected initially with 1 mg/ml Geneticin. During the amplification period, the concentration of Geneticin was gradually reduced to its final concentration of 0.25 mg/ml. For immunoprecipitation experiments, equal number of cells from different lines (FL MOF, MOF  $\Delta$ N, MOF Y416D) along with WT MOCK cells were induced with 100 mM CuSO<sub>4</sub> for one day and harvested. Nuclei were isolated and nuclear extracts were prepared by freeze-thaw cycles in HEMGT 150 buffer (25 mM HEPES 7.6, 0.1 mM EDTA, 12.5 mM MgCl<sub>2</sub>, 10% Glycerol, 0.2% Tween-20 and 150 mM KCl). INPUT

samples were taken from the nuclear extract. Nuclear extracts were incubated with 30 ml bed volume of magnetic Flag Agarose beads (Sigma) for 2 hours at 4°C. The beads were washed four times with HEMGT 150 buffer and eluted with 3xFlag peptide (final 0.4 mg/ml concentration) containing HEMGT 150 buffer overnight. The elutions were precipitated by TCA/Acetone method and resuspended with 4 x Laemmli Buffer.

### **ChIP-Seq Data processing**

ChIP samples were sequenced using Illumina GAIIx machines at the EMBL GeneCore facility. Resulting reads were mapped to the *D. melanogaster* genome (dmel\_r5.11\_FB2008\_08) using the Bowtie software (Langmead et al., 2009) with the following parameters: -n 2 -k 1 --solexa1.3-quals --best. With these settings, the software maps reads only to unique locations in the genome, choosing the best possible hit. The software takes into consideration the read-quality during alignment and removes all reads containing ambiguous nucleotide assignments (N). Reads mapping to heterochromatic regions were discarded from further analysis (Table S1). We divided the *D. melanogaster* genome into non-overlapping 25 bp bins, and counted the number of reads mapped to each bin. These read-counts were then inputted to the DESeq BioConductor package (Anders and Huber, 2010); counts were normalized between IP and input samples by applying a scaling factor accounting for differences in the total numbers of reads per sequencing run. Any bins with read counts of zero were discarded from further analysis. For each bin, DESeq outputted a log2 fold-change (log2FC) value between normalized read counts in the IP and control samples. We used input DNA as the control for MOF-binding and histone H4-binding as the control for the H4K16 acetylation marks. As DNA fragment sizes after sonication was ~200 bp, we smoothed log2FC values using a 400 bp sliding window approach (Kind et al., 2008). The final log2FC values represented the signal from the IP samples relative to controls, with positive values corresponding to enrichments in the IP sample and negative values corresponding to enrichments in control sample. We calculated log2FC thresholds to indicate significant binding or acetylation compared with the control. Since negative log2FC values (i.e. enrichment in control signal) correspond to experimental noise, we fitted a symmetric null-distribution to the density distribution for log2FC values below the mode (PMID: 16732288). We then applied an FDR-adjusted p-value cut-off of 0.05 to identify 25 bp



bins containing significant binding or acetylation. All ChIP-Seq data will be made available at European Nucleotide Archive (ENA) (<http://www.ebi.ac.uk/ena/>).

### **Detection of MOF-bound and H4K16-acetylated genes**

For the classification of binding patterns, gene bodies were defined as all exonic sequences between +500 bp downstream of the transcription start site (TSS) to the 3'-end, as annotated in the Ensembl database. The promoter was defined as the region between -200 bp and the TSS: the region was selected by identifying the mode for average MOF-binding at the 5'-end of all genes which falls about -100 bp upstream of the TSS in both male and female samples and then providing a 100 bp window either side. We excluded any genes that are <600 bp (1,205 loci) in length, as they were too short for this analysis. There is a clear bi-modal distribution dividing partially bound genes from fully bound ones. We used a threshold of 70% (relative to the gene body length) to differentiate between the two types of binding. Next, we classified partially bound genes as promoter bound, if they contained at least one bin with significant binding in the promoter region defined above. We used a similar classification to identify patterns of H4K16 acetylation. In this case, the promoter was defined as the region between the TSS and +500bp downstream, in order to accommodate shift in acetylation patterns towards the interior of genes compared with MOF-binding.

### **Gene expression profiling**

Gene expression was measured using Affymetrix Drosophila2 GeneChips in at least triplicate for wt male and female 3rd instar larva salivary glands. Data analysis was performed using publicly available packages in the BioConductor Software Suite [PMID: 15461798]. Raw .CEL files were processed using GCRMA and probe sets were mapped to genes using annotation available from the Ensembl database (v57) [PMID: 21045057]. Expressed genes were identified as those outputting MAS5.0 'present' calls in all available biological replicates. Microarray data will be available at ArrayExpress upon acceptance of the manuscript [PMID: 21071405].

### **Acknowledgements**

We are grateful to Jan Kadlec for very helpful discussions and for providing Figure S2. We thank Katrin Rheingans for help with artwork. We also thank Thomas Stehle

for preparing nucleosomes. We thank members of the lab for critical reading of the manuscript and helpful discussions. This work was supported by EU funded NoE “EpiGeneSys” awarded to AA. E.H. is a Darwin trust fellow. JMV acknowledges funding from the ESF Exchange Grant program.

## References

- Akhtar, A., and Becker, P.B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in *Drosophila*. *Mol Cell* 5, 367-375.
- Akhtar, A., Zink, D., and Becker, P.B. (2000). Chromodomains are protein-RNA interaction modules. *Nature* 407, 405-409.
- Alekseyenko, A.A., Peng, S., Larschan, E., Gorchakov, A.A., Lee, O.K., Kharchenko, P., McGrath, S.D., Wang, C.I., Mardis, E.R., Park, P.J., *et al.* (2008). A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. *Cell* 134, 599-609.
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol* 11, R106.
- Bachtrog, D. (2008). Positive selection at the binding sites of the male-specific lethal complex involved in dosage compensation in *Drosophila*. *Genetics* 180, 1123-1129.
- Bell, O., Schwaiger, M., Oakeley, E.J., Lienert, F., Beisel, C., Stadler, M.B., and Schubeler, D. (2010a). Accessibility of the *Drosophila* genome discriminates PcG repression, H4K16 acetylation and replication timing. *Nat Struct Mol Biol* 17, 894-900.
- Bell, O., Schwaiger, M., Oakeley, E.J., Lienert, F., Beisel, C., Stadler, M.B., and Schubeler, D. (2010b). Accessibility of the *Drosophila* genome discriminates PcG repression, H4K16 acetylation and replication timing. *Nature structural & molecular biology* 17, 894-900.
- Bell, O., Wirbelauer, C., Hild, M., Scharf, A.N., Schwaiger, M., MacAlpine, D.M., Zilbermann, F., van Leeuwen, F., Bell, S.P., Imhof, A., *et al.* (2007). Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in *Drosophila*. *EMBO J* 26, 4974-4984.
- Bone, J.R., and Kuroda, M.I. (1996). Dosage compensation regulatory proteins and the evolution of sex chromosomes in *Drosophila*. *Genetics* 144, 705-713.
- Buscaino, A., Legube, G., and Akhtar, A. (2006). X-chromosome targeting and dosage compensation are mediated by distinct domains in MSL-3. *EMBO Rep* 7, 531-538.
- Chouard, T. (2011). Structural biology: Breaking the protein rules. *Nature* 471, 151-153.
- Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., *et al.* (2005). Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nature genetics* 37, 391-400.
- Gelbart, M.E., Larschan, E., Peng, S., Park, P.J., and Kuroda, M.I. (2009). *Drosophila* MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nat Struct Mol Biol* 16, 825-832.
- Gu, W., Szauter, P., and Lucchesi, J.C. (1998). Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of *Drosophila melanogaster*. *Dev Genet* 22, 56-64.
- Gupta, A., Guerin-Peyrou, T.G., Sharma, G.G., Park, C., Agarwal, M., Ganju, R.K., Pandita, S., Choi, K., Sukumar, S., Pandita, R.K., *et al.* (2008). The mammalian ortholog of *Drosophila* MOF that acetylates histone H4 lysine 16 is essential for embryogenesis and oncogenesis. *Molecular and cellular biology* 28, 397-409.
- Hallacii, E., and Akhtar, A. (2009). X chromosomal regulation in flies: when less is more. *Chromosome Res* 17, 603-619.
- Hilfiker, A., Hilfiker-Kleiner, D., Pannuti, A., and Lucchesi, J.C. (1997). *mof*, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila*. *EMBO J* 16, 2054-2060.
- Ilik, I., and Akhtar, A. (2009). roX RNAs: Non-coding regulators of the male X chromosome in flies. *RNA Biol* 6.

Kadlec, J., Hallacali, E., Lipp, M., Holz, H., Sanchez-Weatherby, J., Cusack, S., and Akhtar, A. (2011). Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nat Struct Mol Biol* 18, 142-149.

Kim, D., Blus, B.J., Chandra, V., Huang, P., Rastinejad, F., and Khorasanizadeh, S. (2010). Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. *Nat Struct Mol Biol* 17, 1027-1029.

Kimura, A., Umehara, T., and Horikoshi, M. (2002). Chromosomal gradient of histone acetylation established by Sas2p and Sir2p functions as a shield against gene silencing. *Nat Genet* 32, 370-377.

Kind, J., Vaquerizas, J.M., Gebhardt, P., Gentzel, M., Luscombe, N.M., Bertone, P., and Akhtar, A. (2008). Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in *Drosophila*. *Cell* 133, 813-828.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410, 116-120.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.

Larschan, E., Alekseyenko, A.A., Gortchakov, A.A., Peng, S., Li, B., Yang, P., Workman, J.L., Park, P.J., and Kuroda, M.I. (2007). MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol Cell* 28, 121-133.

Meller, V.H., and Rattner, B.P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *EMBO J* 21, 1084-1091.

Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Mueller, J., *et al.* (2006). Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Mol Cell* 21, 811-823.

Moore, S.A., Ferhatoglu, Y., Jia, Y., Al-Jiab, R.A., and Scott, M.J. (2010). Structural and biochemical studies on the chromo-barrel domain of male specific lethal 3 (MSL3) reveal a binding preference for mono- or dimethyllysine 20 on histone H4. *J Biol Chem* 285, 40879-40890.

Morales, V., Straub, T., Neumann, M.F., Mengus, G., Akhtar, A., and Becker, P.B. (2004). Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. *EMBO J* 23, 2258-2268.

Nielsen, P.R., Nietlispach, D., Buscaino, A., Warner, R.J., Akhtar, A., Murzin, A.G., Murzina, N.V., and Laue, E.D. (2005). Structure of the chromo barrel domain from the MOF acetyltransferase. *J Biol Chem* 280, 32326-32331.

Oldfield, C.J., Meng, J., Yang, J.Y., Yang, M.Q., Uversky, V.N., and Dunker, A.K. (2008). Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* 9 Suppl 1, S1.

Park, S.W., Kang, Y., Sypula, J.G., Choi, J., Oh, H., and Park, Y. (2007). An evolutionarily conserved domain of roX2 RNA is sufficient for induction of H4-Lys16 acetylation on the *Drosophila* X chromosome. *Genetics* 177, 1429-1437.

Park, S.W., Kuroda, M.I., and Park, Y. (2008). Regulation of histone H4 Lys16 acetylation by predicted alternative secondary structures in roX noncoding RNAs. *Mol Cell Biol* 28, 4952-4962.

Raja, S.J., Charapitsa, I., Conrad, T., Vaquerizas, J.M., Gebhardt, P., Holz, H., Kadlec, J., Fraterman, S., Luscombe, N.M., and Akhtar, A. (2010). The nonspecific lethal complex is a transcriptional regulator in *Drosophila*. *Mol Cell* 38, 827-841.

Rea, S., Xouri, G., and Akhtar, A. (2007). Males absent on the first (MOF): from flies to humans. *Oncogene* 26, 5385-5394.

Robert, F., Pokholok, D.K., Hannett, N.M., Rinaldi, N.J., Chandy, M., Rolfe, A., Workman, J.L., Gifford, D.K., and Young, R.A. (2004). Global position and recruitment of HATs and HDACs in the yeast genome. *Mol Cell* 16, 199-209.

Rodriguez, M.A., Vermaak, D., Bayes, J.J., and Malik, H.S. (2007). Species-specific positive selection of the male-specific lethal complex that participates in dosage compensation in *Drosophila*. *Proc Natl Acad Sci U S A* 104, 15412-15417.

Sanjuan, R., and Marin, I. (2001). Tracing the origin of the compensasome: evolutionary history of DEAH helicase and MYST acetyltransferase gene families. *Mol Biol Evol* 18, 330-343.

Schwaiger, M., Stadler, M.B., Bell, O., Kohler, H., Oakeley, E.J., and Schubeler, D. (2009). Chromatin state marks cell-type- and gender-specific replication of the *Drosophila* genome. *Genes Dev* 23, 589-601.

Shimojo, H., Sano, N., Moriwaki, Y., Okuda, M., Horikoshi, M., and Nishimura, Y. (2008). Novel structural and functional mode of a knot essential for RNA binding activity of the Esa1 presumed chromodomain. *J Mol Biol* 378, 987-1001.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311, 844-847.

Smith, E.R., Cayrou, C., Huang, R., Lane, W.S., Cote, J., and Lucchesi, J.C. (2005). A human protein complex homologous to the *Drosophila* MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol Cell Biol* 25, 9175-9188.

Straub, T., and Becker, P.B. (2007). Dosage compensation: the beginning and end of generalization. *Nat Rev Genet* 8, 47-57.

Suganuma, T., Gutierrez, J.L., Li, B., Florens, L., Swanson, S.K., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2008). ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. *Nat Struct Mol Biol* 15, 364-372.

Suka, N., Luo, K., and Grunstein, M. (2002). Sir2p and Sas2p opposingly regulate acetylation of yeast histone H4 lysine16 and spreading of heterochromatin. *Nat Genet* 32, 378-383.

Sural, T.H., Peng, S., Li, B., Workman, J.L., Park, P.J., and Kuroda, M.I. (2008). The MSL3 chromodomain directs a key targeting step for dosage compensation of the *Drosophila melanogaster* X chromosome. *Nat Struct Mol Biol* 15, 1318-1325.

Taipale, M., Rea, S., Richter, K., Vilar, A., Lichter, P., Imhof, A., and Akhtar, A. (2005). hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol Cell Biol* 25, 6798-6810.

Thomas, T., Dixon, M.P., Kueh, A.J., and Voss, A.K. (2008). Mof (MYST1 or KAT8) is essential for progression of embryonic development past the blastocyst stage and required for normal chromatin architecture. *Molecular and cellular biology* 28, 5093-5105.

Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* 138, 1019-1031.

Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., and Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* 138, 1122-1136.

## Figure Legends

### Figure 1. MOF is the major H4K16 specific HAT in the male and female genome.

(A) H4K16ac is lost from all chromosomes in the absence of MOF. Immunostaining of polytene chromosomes from male and female third instar larva salivary glands, using antibodies against H4K16ac, MSL1 and HA (MOF) as indicated. DNA staining is shown in blue (Hoechst 322). (B) Western blot analysis of extracts prepared from wt, *mof2* and FL-MOF expressing male and female 3<sup>rd</sup> instar larva showing the effect of MOF depletion on H4K16ac. (C) Number of eclosed *mof2* adults compared to the control carrying the wt MOF allele. While males show full lethality, female number is reduced approx. two fold. Error bars represent the standard deviation from three independent experiments. (D) Lifespan of control, *mof2*, and FL-MOF expressing female flies after eclosion. Error bars represent the standard error of the mean.

### Figure 2. The chromobarrel domain is essential for MOF function.

(A) Male lethality upon disruption of the chromobarrel domain. Male viability was assayed upon expression of various MOF transgenes in the *mof2* background as indicated. Percentages refer to the number of *mof2* males compared to the number of heterozygous *mof2* females resulting from the same cross. (B-E) Immunostaining of polytene chromosomes from *mof2* male third instar larvae salivary glands, expressing FL (B),  $\Delta$ Zn (C), Y416D and W416G (D), and  $\Delta$ N MOF (E) transgenes. Antibodies against HA (MOF), MSL1 and H4K16ac were used as indicated in the figure. DNA staining is shown in blue (Hoechst 322).

### Figure 3. The MOF chromobarrel domain is required for MSL spreading.

(A) MSL targeting is compromised without the MOF N-terminus. ChIP using MSL1 and MSL3 antibodies in wt male 3<sup>rd</sup> instar larva as well as *mof2* larva, or *mof2* larva that express FL MOF or  $\Delta$ N MOF transgenes. Binding to the X-linked high affinity sites at the *roX2* gene and location 15A8, as well as the X chromosomal genes *Rpl22*, *Klp3a*, and *Ucp4a* is shown. PKA is used as a negative control. Primers were positioned at the promoter (P), middle (M), and end (E) of genes. The exact position of the primers is described in the Supplemental Data. ChIP is shown as percentage recovery of input DNA (% Input). Error bars represent standard deviation (StDev) of three independent experiments. (B) MSL spreading is lost upon disruption of the

chromobarrel domain. Same as in (A), using *mof2* larvae that express FL MOF or Y416D MOF, respectively.

**Figure 4. The chromobarrel domain is required for H4K16ac.**

(A) Compromised MOF targeting of  $\Delta N$  MOF. ChIP using MOF (upper panel), and H4K16ac (lower panel) antibodies in wt male 3<sup>rd</sup> instar larva as well as *mof2* larva or *mof2* larva that express FL-MOF or  $\Delta N$  MOF transgenes. Binding to the autosomal genes *bt*, *cg3937*, *cg6729*, *cg6884*, *cg31866*, *cg2709*, *cg7638*, the X-linked high affinity sites at the *roX2* gene and location 15A8, as well as the X chromosomal genes *Rpl22*, *Klp3a*, and *Ucp4a* is shown. Primers were positioned at the promoter (P), middle (M), and end (E) of genes. The exact position of the primers is described in the Supplemental Data. ChIP is shown as percentage recovery of input DNA (% Input). Error bars represent standard deviation (StDev) of three independent experiments. (B) H4K16ac is lost upon disruption of the chromobarrel domain, while promoter binding of MOF is retained. Same as in (A), using *mof2* larvae that express FL MOF or Y416D MOF, respectively.

**Figure 5. The MOF chromobarrel domain functions after chromatin binding.**

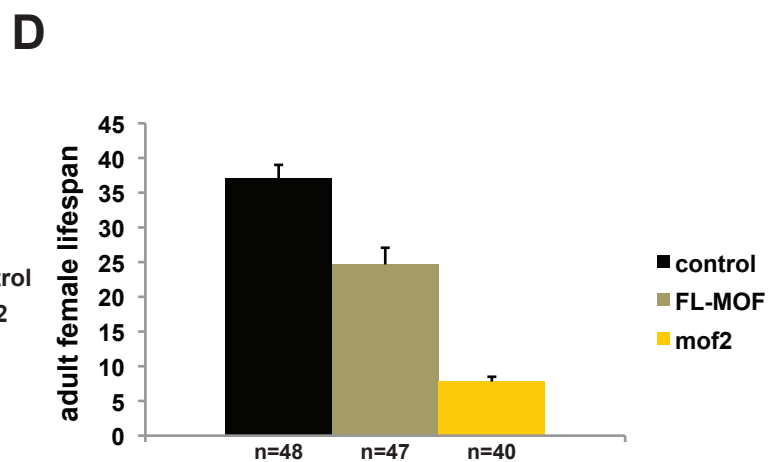
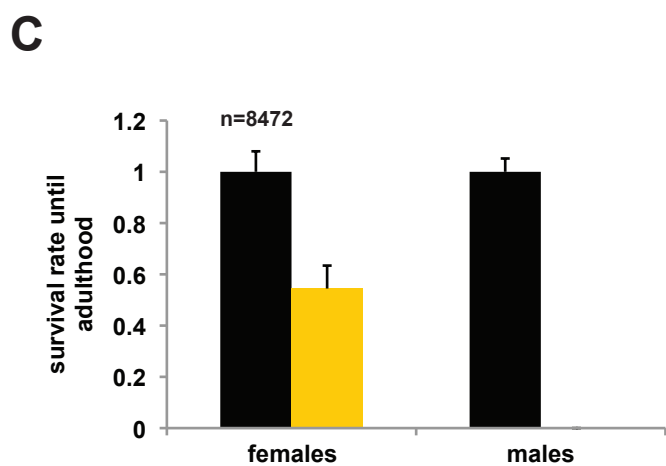
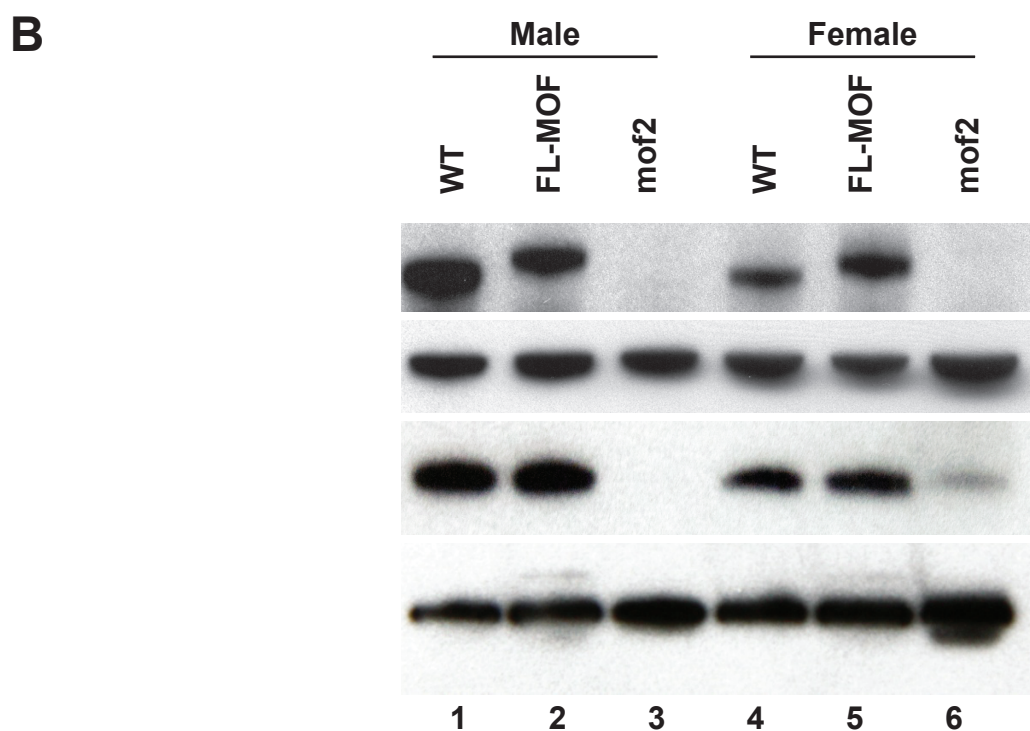
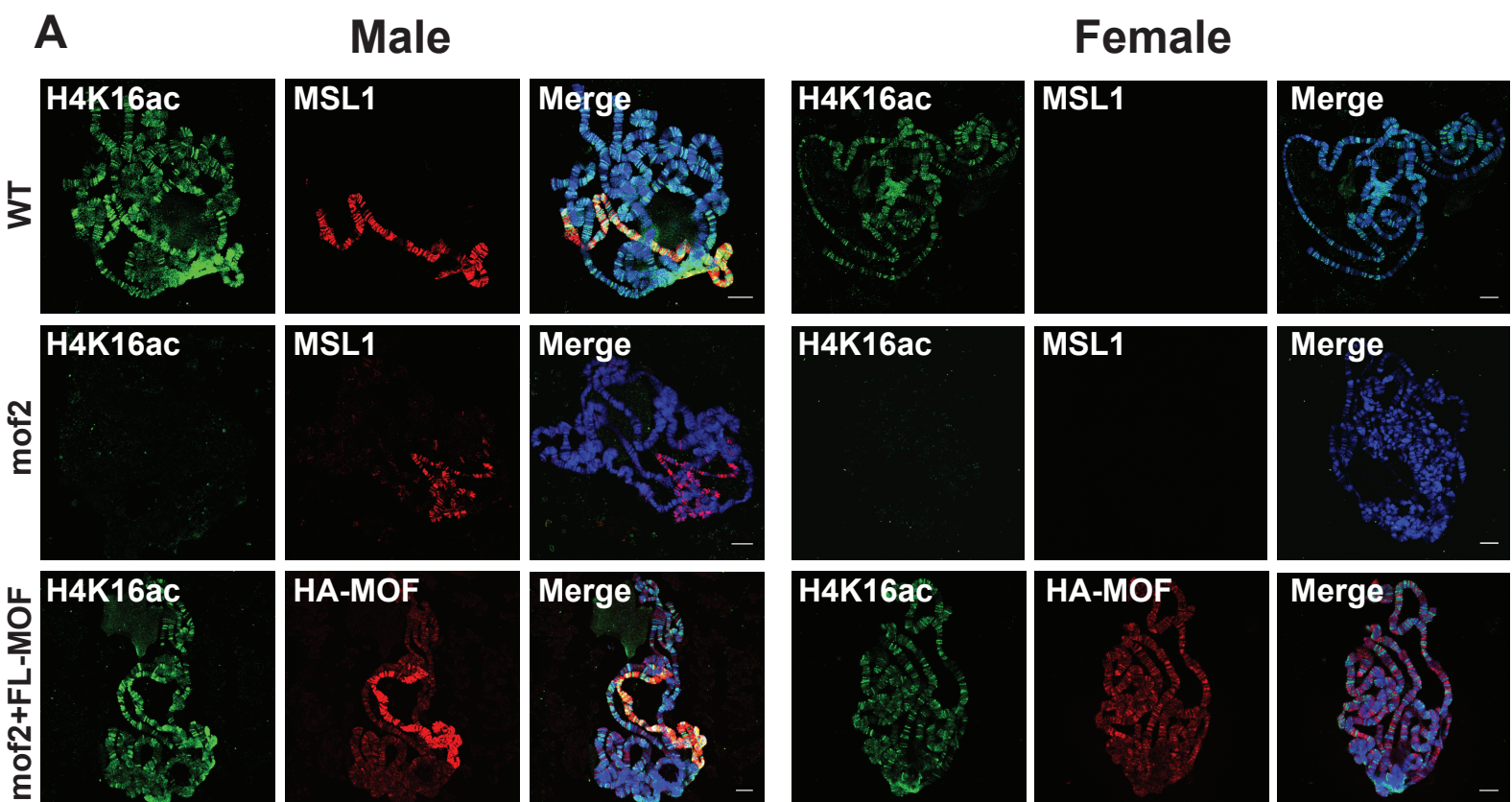
(A) Bulk H4K16ac is lost upon disruption of the chromobarrel domain. Western blot analysis of extracts from *mof2* male 3<sup>rd</sup> instar larva expressing the indicated MOF transgenes. (B) Acetylation assay on native nucleosomes. Each reaction contains 5.3 nM of FL, Y416D or W426G MOF proteins, with or without MSL1 and MSL3, respectively, and 1.5  $\mu$ g of native nucleosomes purified from MCF-7 cells. (C) Same as in (B), using FL and  $\Delta N$  MOF proteins. (D) Acetylation assay on free histone octamers. 40 ng (21.3 nM) of the respective recombinant MOF-protein or trimeric complexes were incubated with C<sup>14</sup> labeled acetyl coenzyme A and 1.5  $\mu$ g of recombinant (*Xenopus*) histone octamer, as indicated in the figure. The reaction was then applied on a P81 filter paper, air dried, washed and counted in scintillation liquid. (E) Disruption of the chromobarrel domain does not affect nucleosomal binding. Nucleosome pull down experiments are shown, using trimeric complexes containing FL, Y416D, W426G or  $\Delta N$  MOF together with MSL1 and MSL3 proteins. Purified endogenous histone octamers from *Drosophila* embryos were wrapped with 147 bp biotinylated DNA to assemble mononucleosomes and coupled to streptavidin beads. After incubation with trimeric complexes and stringent washing, MOF binding

was detected by western blot analysis using HA antibodies. 5% of the Input material (lanes 1-4), IP using empty streptavidin beads as a control (mock-IP) (lanes 5-8, and nucleosome IP (nuc-IP) (lanes 9-12) are shown.

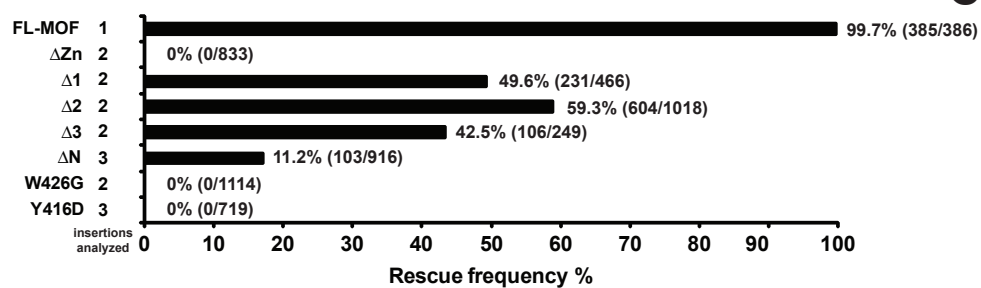
**Figure 6. Several layers of regulation control genomewide H4K16ac in *Drosophila*.**

Our data suggests that in *Drosophila* MOF is responsible for the majority of H4K16ac at gene promoters in males and females, and for hyperacetylation along the male X chromosome. MOF activity is tightly regulated to achieve these diverse tasks. The N-terminal region is required for targeting of MOF to gene promoters and for assembly the MSL complex. At the same time the N-terminus constrains the enzymatic activity of the MOF HAT domain, which is only unleashed in the context of the MSL complex to achieve hyperacetylation of the male X (blue arrows). The next layer of regulation is imposed by the chromobarrel domain (CD). Following binding of MOF to its chromatin target sites, nucleic acid interactions of the chromobarrel domain are required to trigger acetylation of the H4 tail (red arrows), which subsequently allows MSL complex spreading along the transcribed regions of genes. The chromobarrel domain thus acts as an “on-off switch” for H4K16ac.

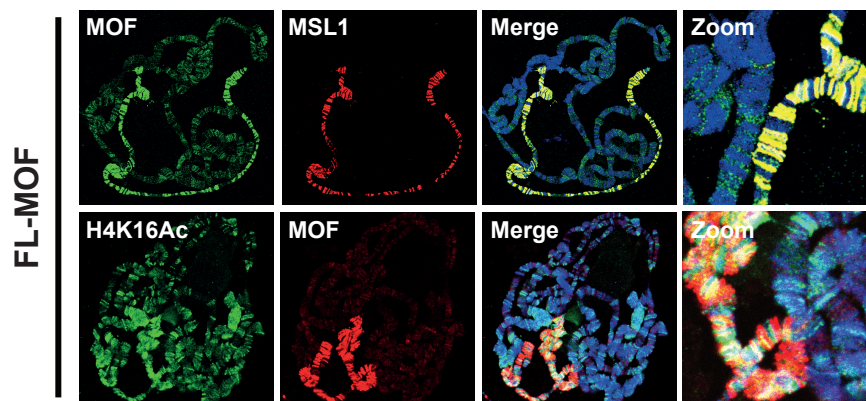




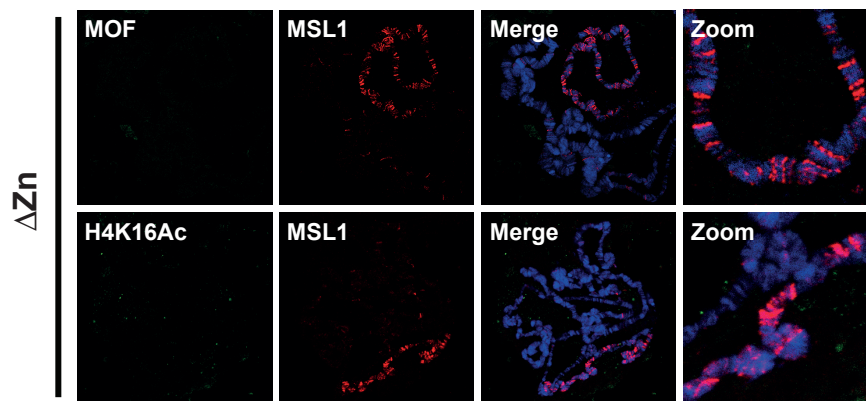
A



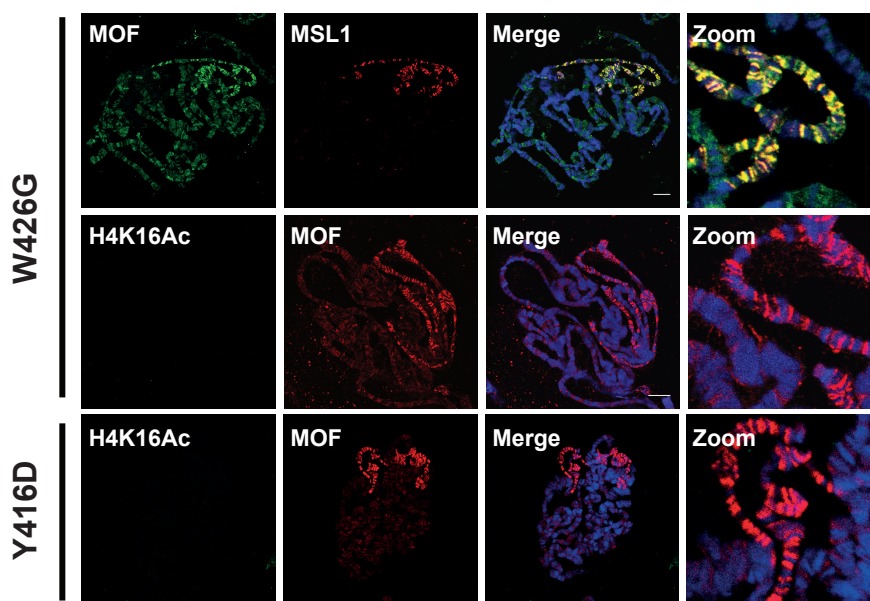
B



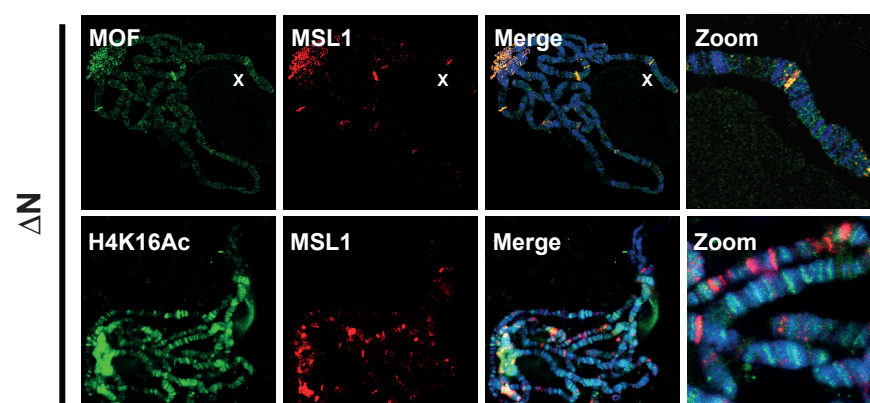
C



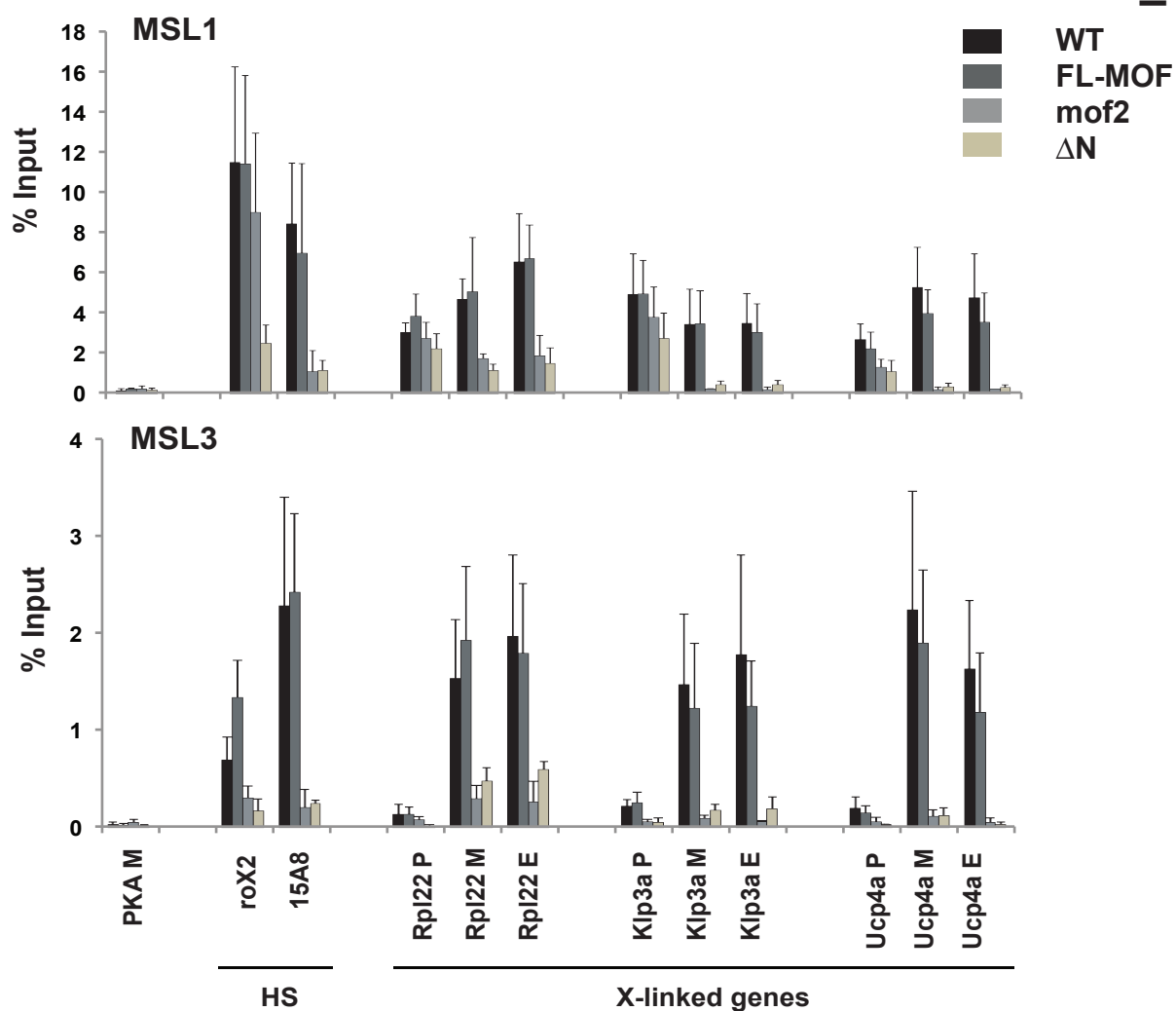
D



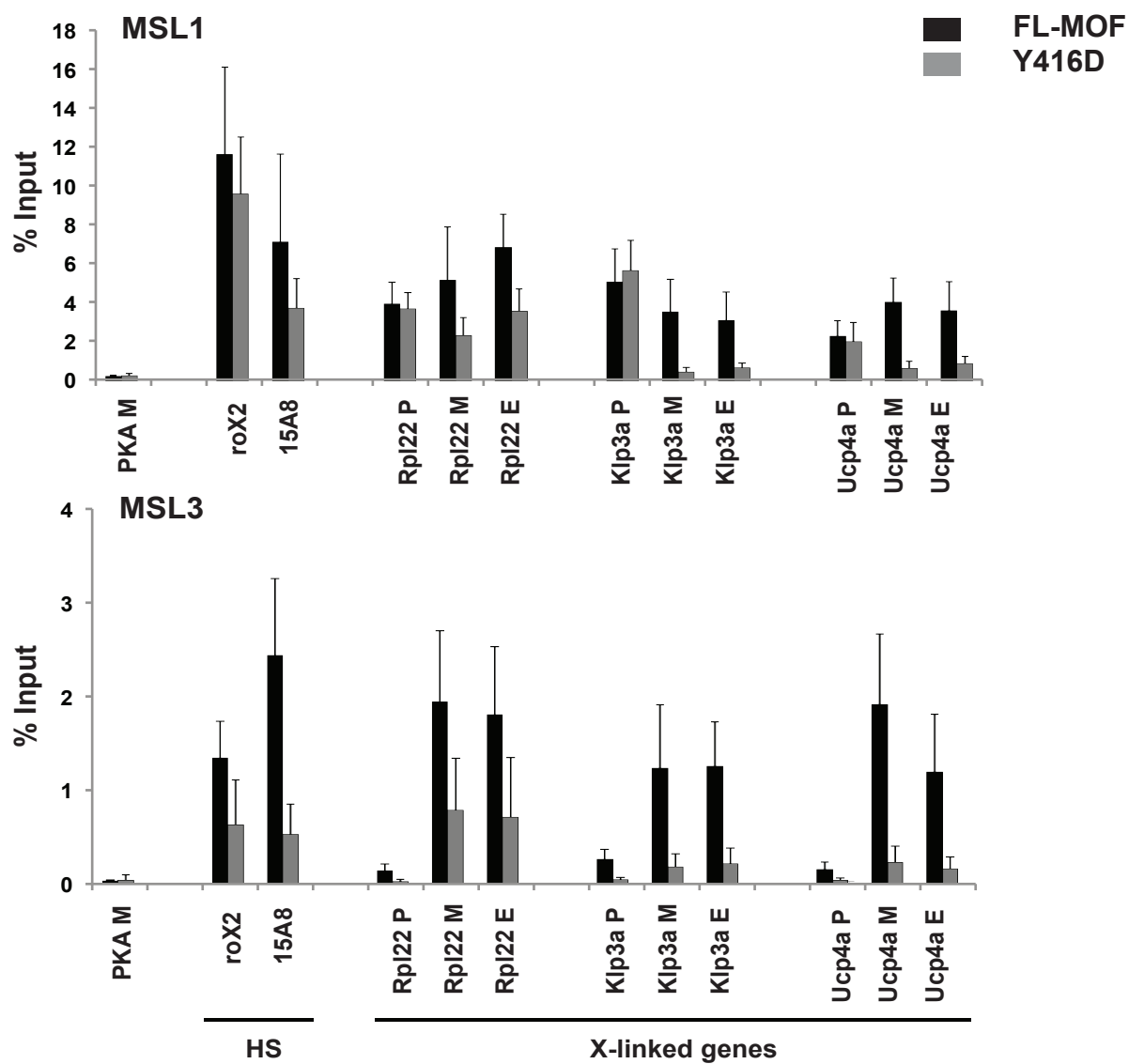
E



A



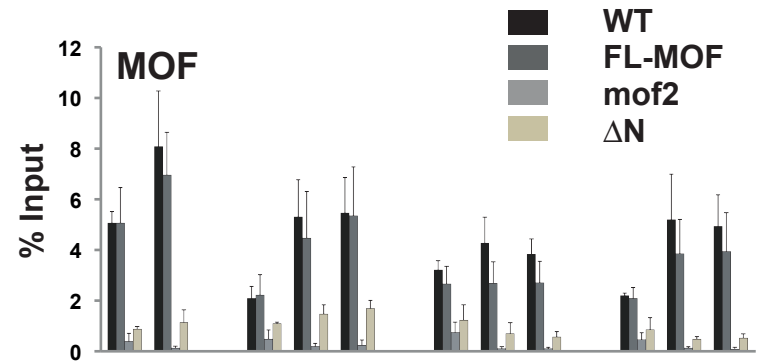
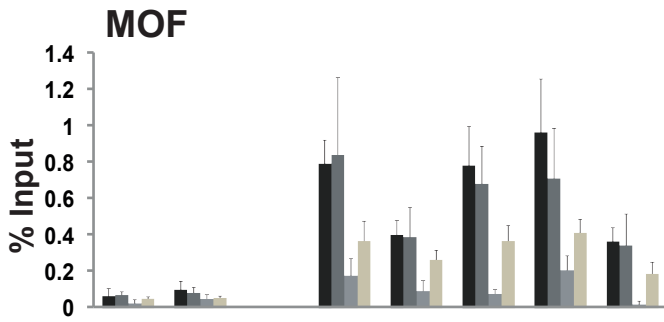
B



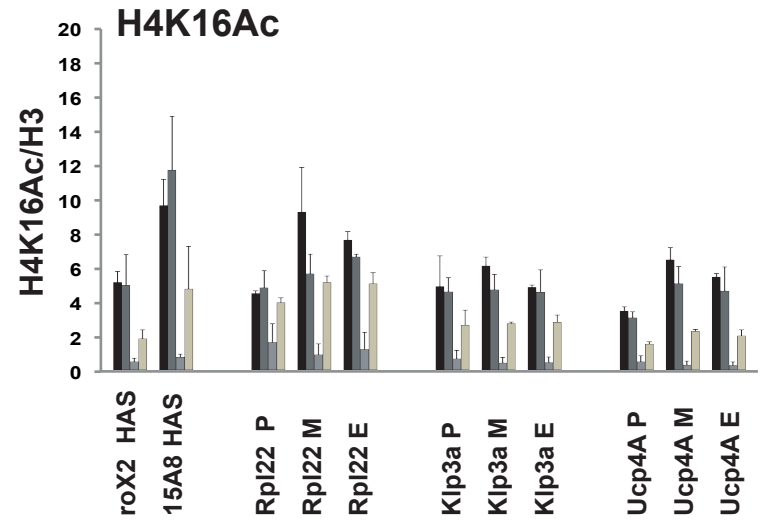
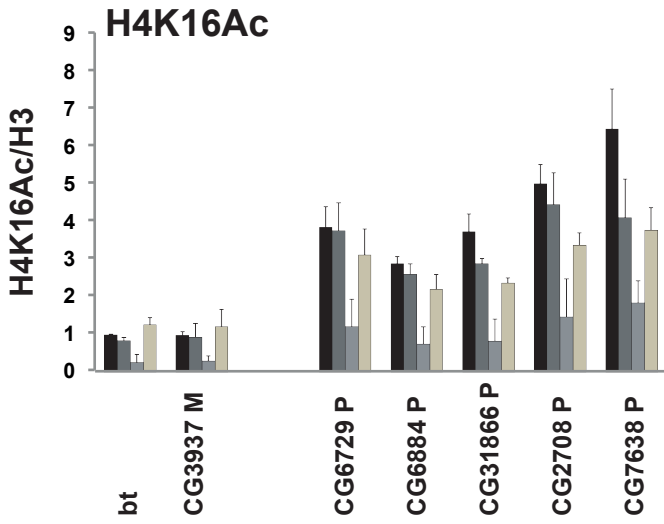
Autosomes

X chromosome

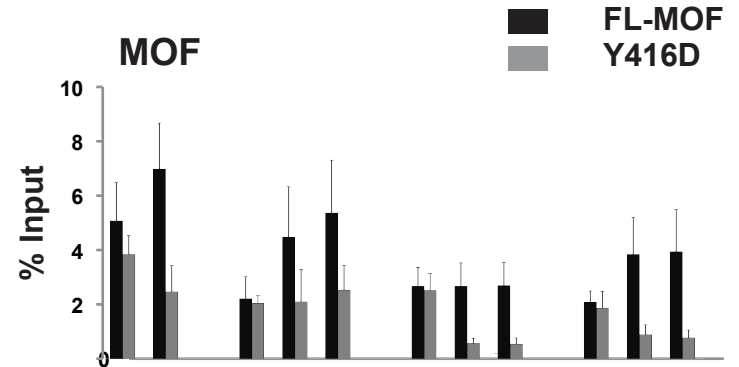
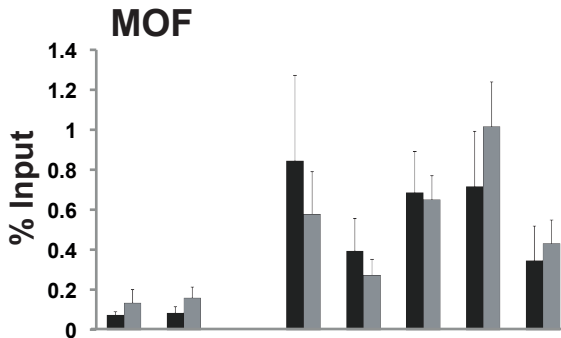
A



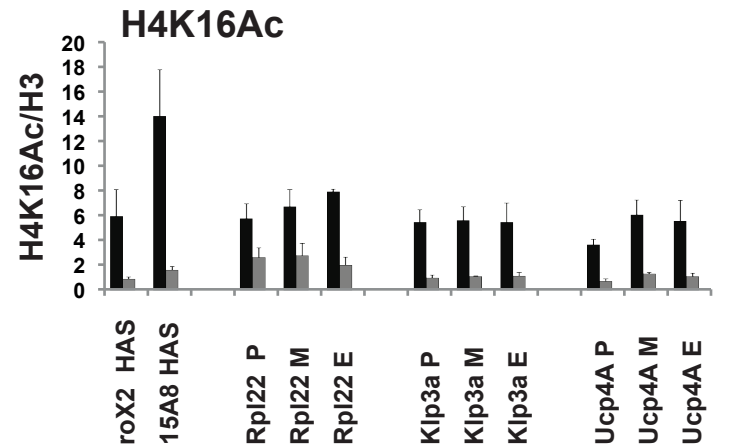
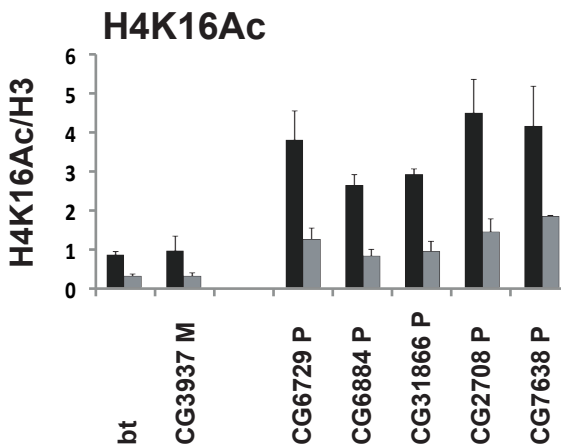
B



C

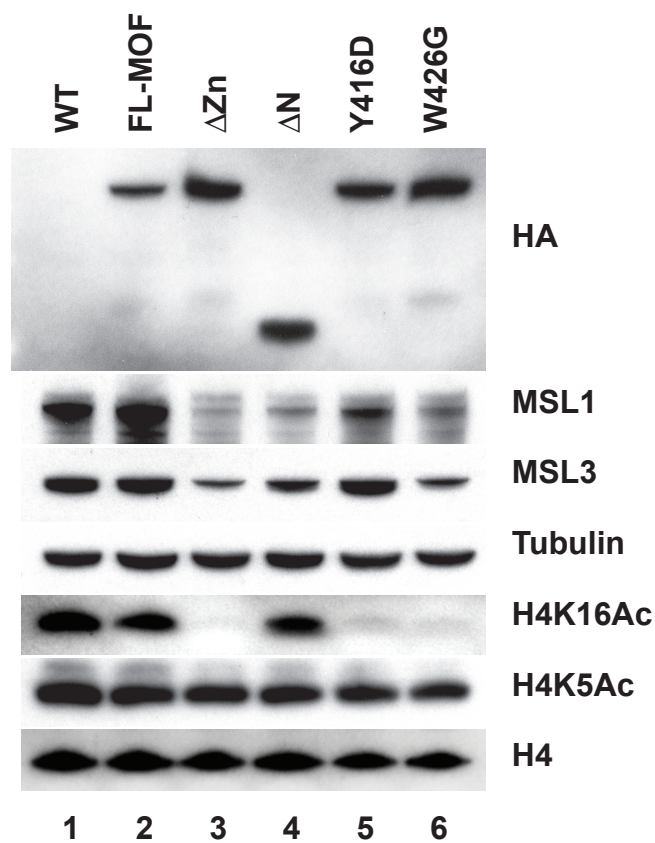


D

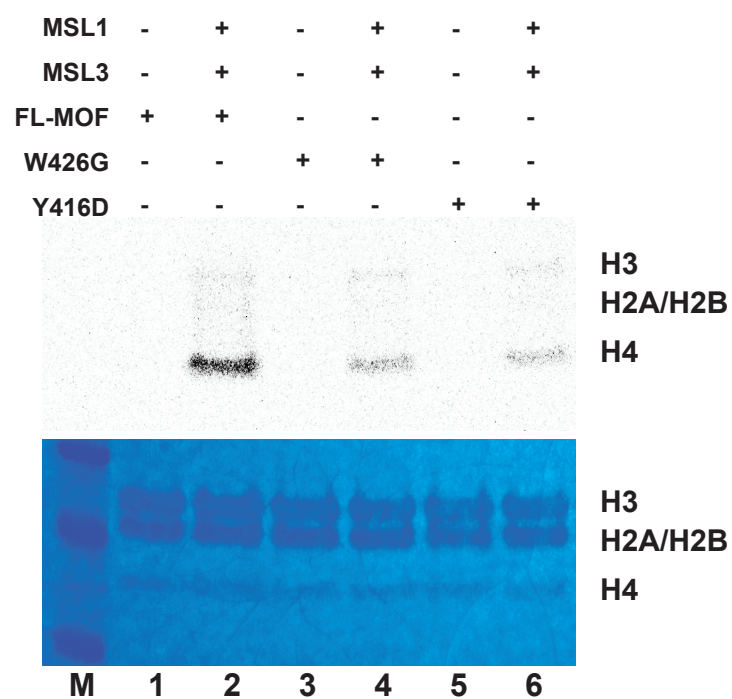




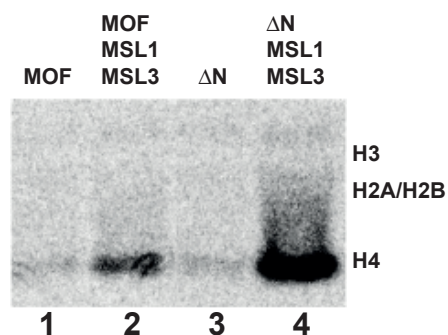
**A**



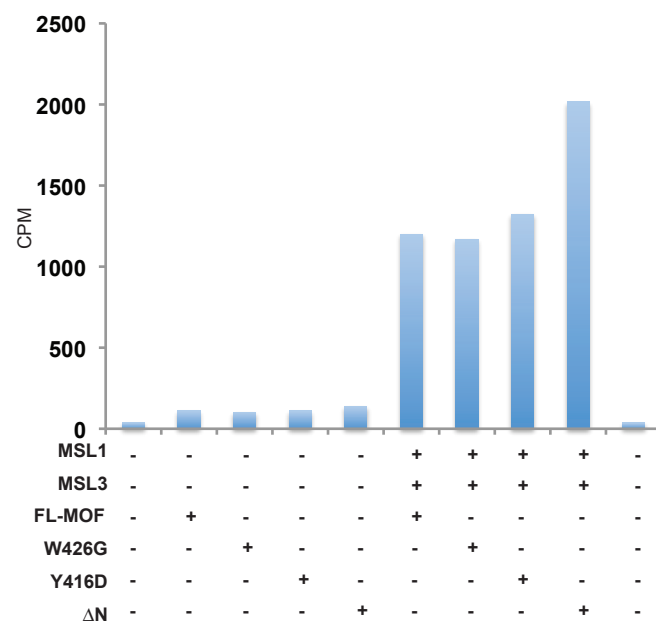
**B**



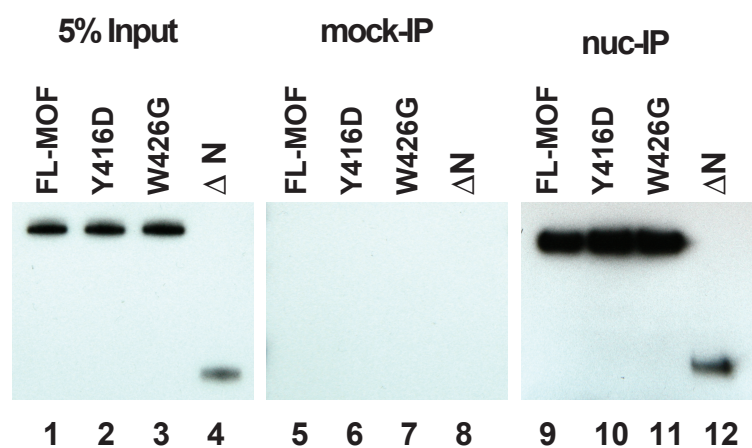
**C**

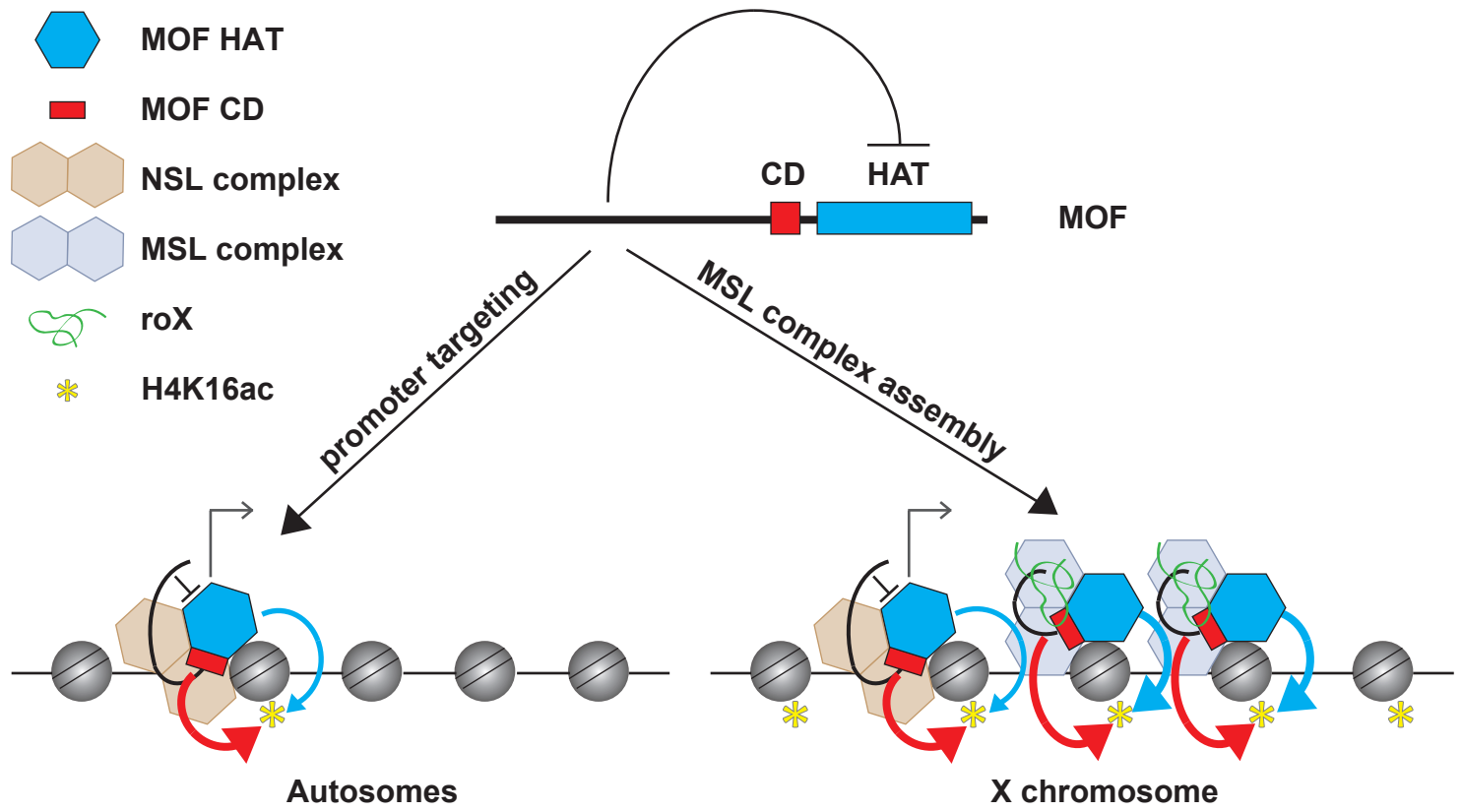


**D**



**E**





## Supplementary Information

### The MOF chromobarrel domain controls genomewide H4K16ac and spreading of the MSL complex

Thomas Conrad<sup>1,2</sup>, Florence Cavalli<sup>3</sup>, Erinc Hallacli<sup>1,2</sup>, Herbert Holz<sup>1</sup>, Jop Kind<sup>2,4</sup>, Ibrahim Ilik<sup>1,2</sup>, Juan M. Vaquerizas<sup>3</sup>, Nicholas Luscombe<sup>3</sup> and Asifa Akhtar<sup>1,\*</sup>

#### Supplementary Methods

##### Primers used in this study

##### ChIP primers:

F=Forward, R=Reverse, P=promoter, m=middle, 3'= towards the 3' end of the gene.

UCP4a P F:	CAA GTT GTC GCG AGT TGA AA
UCP4a P R:	CAA TTG CTT CGC TCT AGC TG
UCP4a m F:	CGC AAG GAG TTC ACA CAG AA
UCP4a m R:	CTC CAT TTG GAT TTG CAC CT
UCP4a 3' F:	TTC ATG TTA CCC CGC CTT TA
UCP4a 3' R:	CTC CTG ACA TTT GGG CAT TC
Rpl22 P F:	CAA TCC AAT GCG CAG TTA TG
Rpl22 P R:	AAG GCC TTG TTC GCA TAT TG
Rpl22 m F:	TAG CGG TAA GCT GGG CTA AA
Rpl22 m R:	GTC GCT CTG ATG GCA GTG TA
Rpl22 3' F:	GGC TAG CCC GAA GTT TTC TT
Rpl22 3' R:	AGC TGA TCC CTT CAG TGG AA
Klp3a P F:	TTG GCT CTT GAA TAG CAG ATT T
Klp3a P R:	GTG CAC AAA TCG TCC AAC C
Klp3a m F:	GCA GCT CCT GTT TGA GAT CC
Klp3a m R:	CAT TCC CAT TCG GAG GAG TA
Klp3a 3' F:	ATG GGC TGT CAG GCA TTT AG
Klp3a 3' R:	GAG GAG CAG CAA AAG AGG TG
15A8 F:	TGA CGC CTT TGC TGA ATG T

15A8 R:	TGC GCT CCT ATC ACC CAG A
roX 3' F:	ACG GTG CTG GCT TAG AGA GA
roX 3' R:	GGC GGA AAT GTA TTT GCA GT
CG6729 P F:	TATCCGGGTAAGTCCAGGTG
CG6729 P R:	CAATTTCCAGCGAAAATGTTG
CG31866 P F:	CGCAAACAGTTGGATTTTCC
CG31866 P R:	GGCCAGAACTCTGTCTGGTAA
CG6884 P F:	TGGGACGGAACCTTGCTTATC
CG6884 P R:	GGCTTGTCTCTGGAAAGATCA
CG2708 P F:	GCAGCCGGTTCAATAGTCTC
CG2708 P R:	CGATCGCTCTCGAATACACA
CG7638 P F:	ACACATTTTCTCGGCTCCAC
CG7638 P R:	TTTTCCGTCATGTCCACTCA
CG3937 m F:	AATCTTGGCATCCAGCTCAC
CG3937 m R:	AATTCATGGTCCGCAAGAAC

**Expression primers:**

roX2 exon3 F:	TCGCAATGCAAACCTGAAGTC
roX2 exon3 R:	AGGCGCGTAAAACGTTACC



## Supplementary Figure Legends

### Figure S1. MOF acetylates promoters across the male and female genome

(A) Genomic patterns of MOF-binding and H4K16 acetylation. Screenshots of the Integrated Genome Browser show the patterns (log2FC values) of MOF-binding (dark colour) and H4K16 acetylation (light colour) in male (blue colour shades) and female (red colour shades) salivary glands along a 20kb section of chromosome 3L (left) and a 25 kb section of the X chromosome (right). Genomic coordinates and the locations of annotated Refseq genes on the forward (+) and reverse (-) strands are indicated below. (B) Profiles of MOF-binding and H4K16 acetylation along genes. The plots show the average log2FC signal for MOF-binding (red) and H4K16ac (black) along genes in wt males (left) and females (right). Log2FC signals are shown for the region surrounding the TSS (-2000bp to +2000bp) and for the 3'-end of the gene (3'-end to -2000bp upstream of the transcription termination site). Average log2FC values for X-linked and autosomal genes are represented by solid and dashed lines, respectively. (C) Overlap in MOF-bound (left) and H4K16-acetylated (right) autosomal (top) and X-linked genes (bottom) genes between males and females. (D) Overlap of MOF-bound with H4K16-acetylated genes in wt males and females. Autosomal (left) and X-linked (right) genes were divided into three categories showing full (full), promoter-only (prom) or no (not) binding by MOF. For each category, the stacked barplot specifies the number of genes showing full (full), promoter-only (prom) or no (not) H4K16 acetylation. The numbers within each block indicate the percentage of genes within this block relative to all genes in the category. (E) Overlap of MOF-bound and H4K16-acetylated genes detected by the ChIP-chip analyses performed by Kind et al. (2008) from cell lines and our ChIP-seq analysis using third instar salivary gland samples. Both male and female samples are presented.

### Figure S2. Analysis of the MOF chromobarrel domain

(A-C) MOF chromo-barrel domain structure. Ribbon diagram of the *Drosophila* MOF chromo-barrel domain (A). Residues Tyr416 and Trp426 are shown as sticks. Structure of the yeast Esa1 chromodomain (residues 17-89, B) in the same orientation as MOF. Structure of the extended chromodomain of yeast Esa1 (residues 1-89, C). The N- and C- termini form a short  $\beta$ -sheet and a loop following Tyr56 changes its conformation. (D-F) Conservation of the chromobarrel domain

surfaces in MOF and MSL3. (D) Ribbon representation of the complex of MSL3, DNA and the H4 peptide (PDB code 3OA6). The structure of the MOF chromobarrel domain (PDB code 2BUD) is superimposed onto the one of MSL3. (E) Surface representation of the MSL3 chromobarrel domain in the same orientation as in (D) highlighting areas of conserved residues involved in DNA and H4 binding. The conservation of the surface is represented from gray to red (red is 100 % conserved) according to the color scale bar. The aromatic cage around Phe56 includes also Tyr31, Trp59 and Trp63. (F) Surface conservation of the MOF chromobarrel domain based on ortholog sequences ranging from human to nematodes. The conserved surface corresponding to the MSL3 DNA binding region includes Asn420, Arg422, Tyr314, Trp426 and Arg387. Tyr416 forms a floor of a cavity surrounded by Leu419, Leu423 and His393.

### **Figure S3. MOF transgenes analyzed in this study**

(A) Schematic representation of the domain structure of the known MOF alleles *mof1* and *mof2*, as well as the MOF derivatives generated for this study. Globular domains in the MOF protein are the chromobarrel domain (chromo), the C-terminal HAT domain (HAT) containing a zinc finger (Zn). G691 is located in the catalytic center of the MOF HAT domain. Y416D and W426G reside in the chromobarrel domain and disrupt RNA binding. (B-I) MOF transgenes are expressed to wt levels. Western blot analysis using extracts from male adult fly heads expressing the indicated transgenes. Two to three independent lines were analyzed for each mutant variant. Tubulin (Tub) was used a loading control.

### **Figure S4. Nuclear localization of the $\Delta$ Zn MOF protein, and conservation of MOF**

Immunostaining of salivary glands from *mof2* male third instar larvae expressing the FL-MOF or  $\Delta$ Zn MOF transgene, using antibodies against HA (MOF) and MSL1 as indicated. DNA staining is shown in blue (Hoechst 322).

### **Figure S5. roX RNA levels are reduced upon removal of the MOF N-terminus**

(A) roX2 levels in *mof2* male larva salivary glands expressing MOF transgenes as indicated. Transcript levels were normalized to roX2 DNA recovered from the same sample. (B) Alignment of the MOF amino acid sequence in different animal species.

Evolutionary distance between *Drosophila melanogaster* and *virilis* is approx. 55 mio years, *melanogaster* and *sechellia* approx. 2,5 mio years. Sequence conservation is indicated by darker greyscales.

**Figure S6. MOF chromatin targeting is lost upon deletion of the zinc finger**

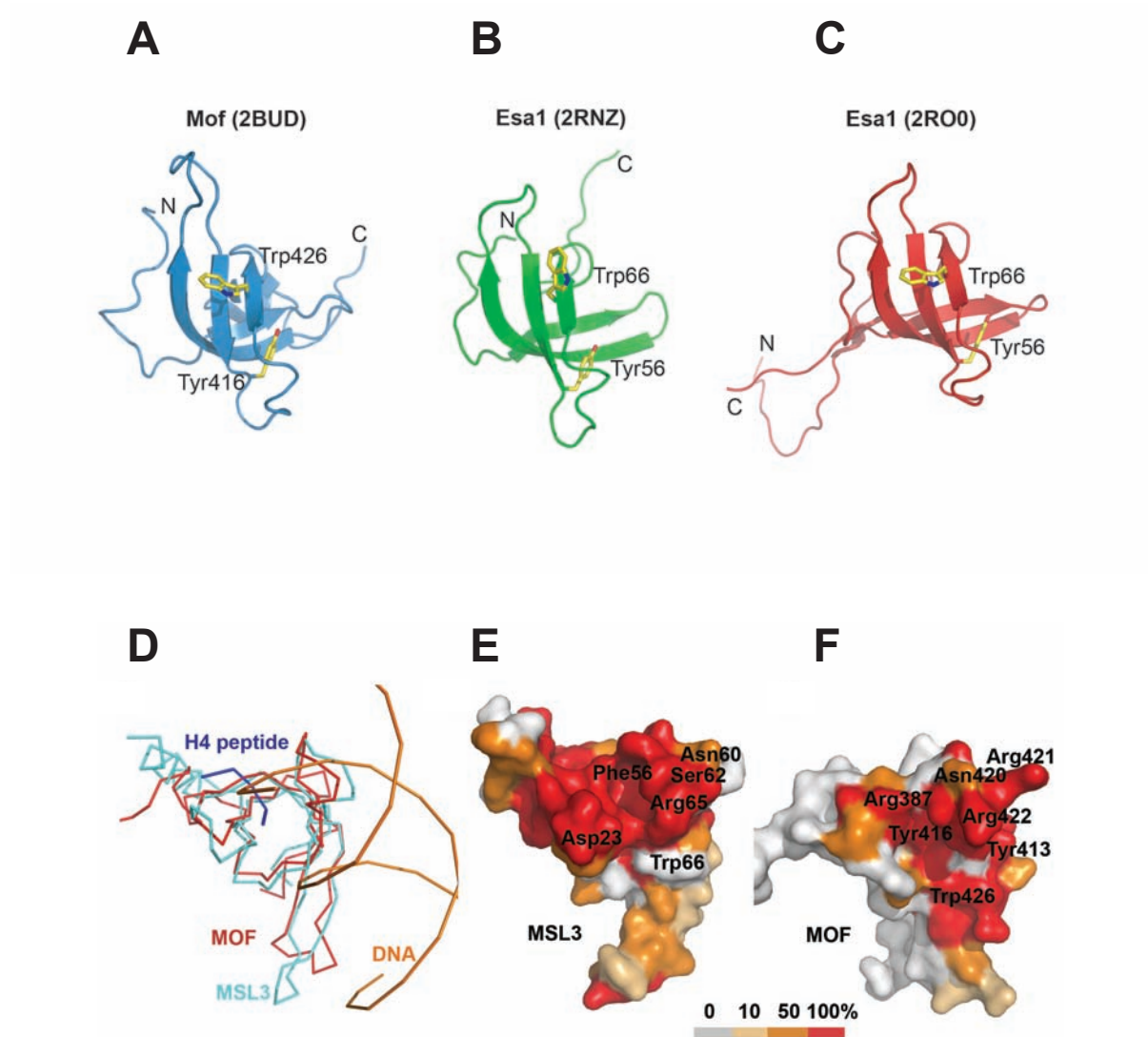
(A) MSL targeting is compromised upon deletion of the MOF zinc finger. ChIP using MSL1 (upper panel) and MSL3 antibodies (lower panel) in wt male 3<sup>rd</sup> instar larva as well as *mof2* larva, or *mof2* larva that express FL MOF or  $\Delta$ Zn MOF transgenes. Binding to the X-linked high affinity sites at the *roX2* gene and location 15A8, as well as the X chromosomal genes *Rpl22*, *Klp3a*, and *Ucp4a* is shown. PKA is used as a negative control. Primers were positioned at the promoter (P), middle (M), and end (E) of genes. (B) Compromised MOF targeting of  $\Delta$ Zn MOF. ChIP using MOF (upper panel), and H4K16ac (lower panel) antibodies in wt male 3<sup>rd</sup> instar larva as well as *mof2* larva or *mof2* larva that express FL-MOF or  $\Delta$ N MOF transgenes. Binding to the autosomal genes *bt*, *cg3937*, *cg6729*, *cg6884*, *cg31866*, *cg2709*, *cg7638*, the X-linked high affinity sites at the *roX2* gene and location 15A8, as well as the X chromosomal genes *Rpl22*, *Klp3a*, and *Ucp4a* is shown. The exact position of the primers is described in the Supplemental Data. ChIP is shown as percentage recovery of input DNA (% Input). Error bars represent standard deviation (StDev) of three independent experiments.

**Figure S7. The MOF N-terminus is required for incorporation into the NSL complex**

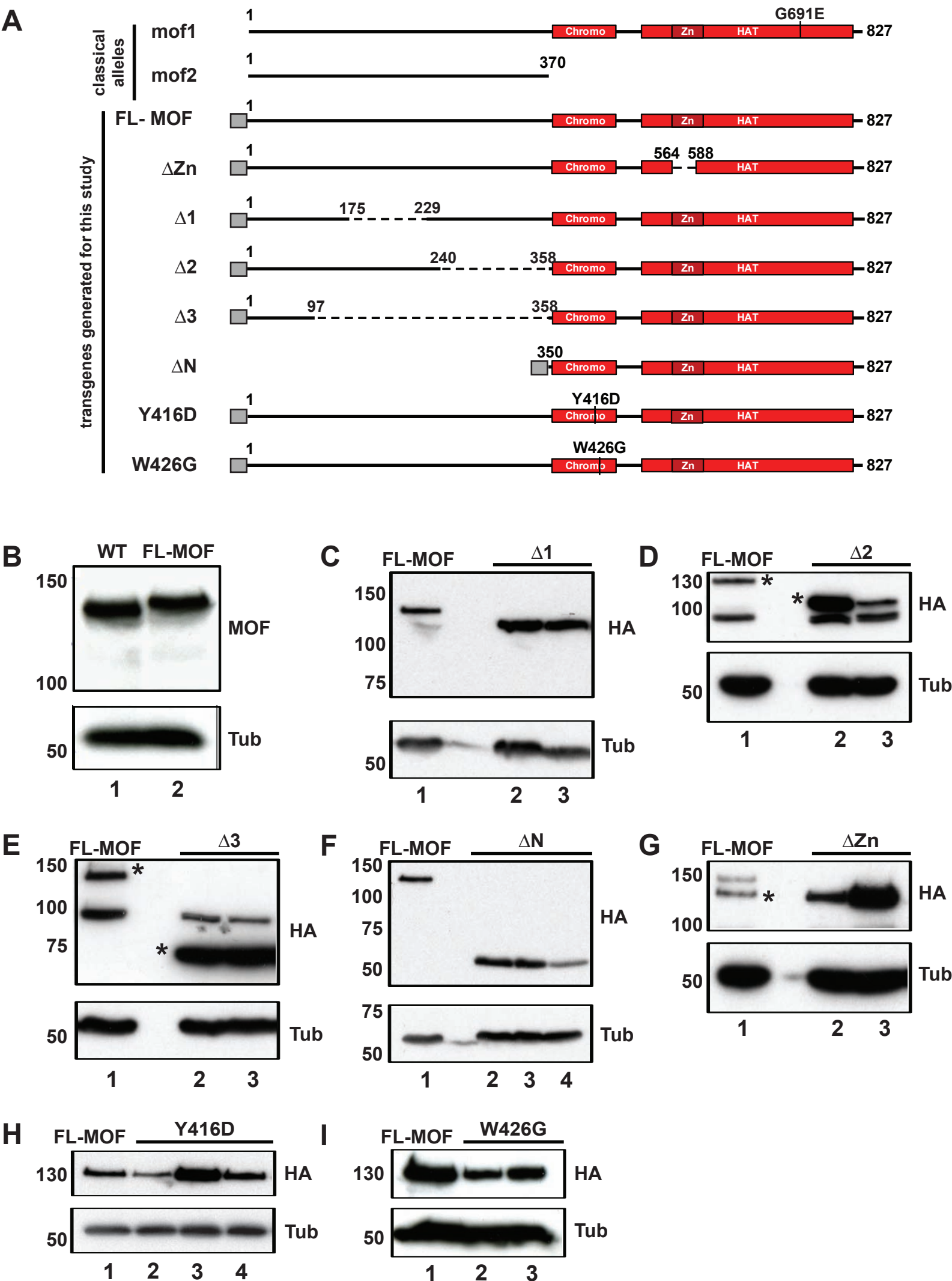
Coimmunoprecipitations were performed from nuclear extracts obtained from SL-2 cell lines stably expressing FLAG-tagged wt, Y416D and  $\Delta$ N MOF. Input material and immunocomplexes were subjected to western blot analysis using the indicated antibodies for detection.



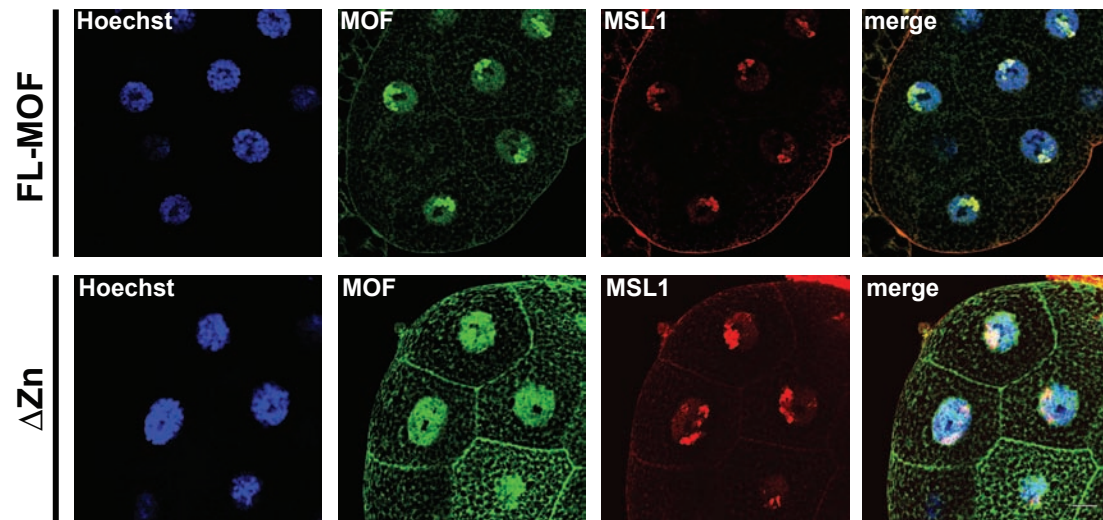
# Supplementary Figure 2



# Supplementary Figure 3

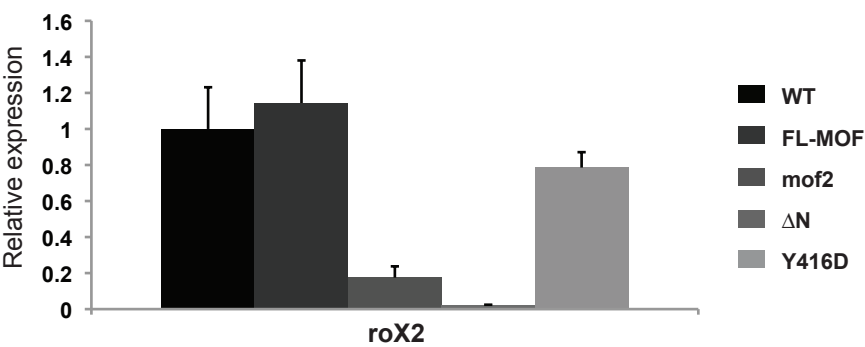


# Supplementary Figure 4

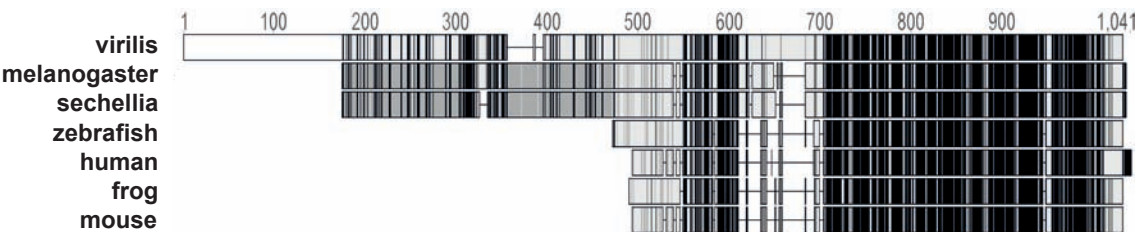


# Supplementary Figure 5

A



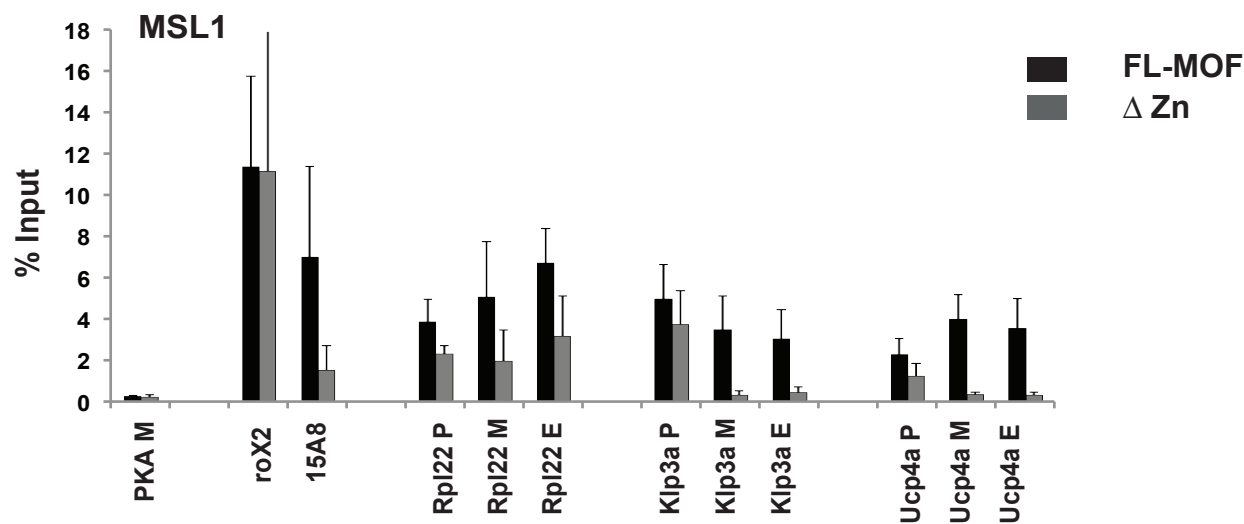
B



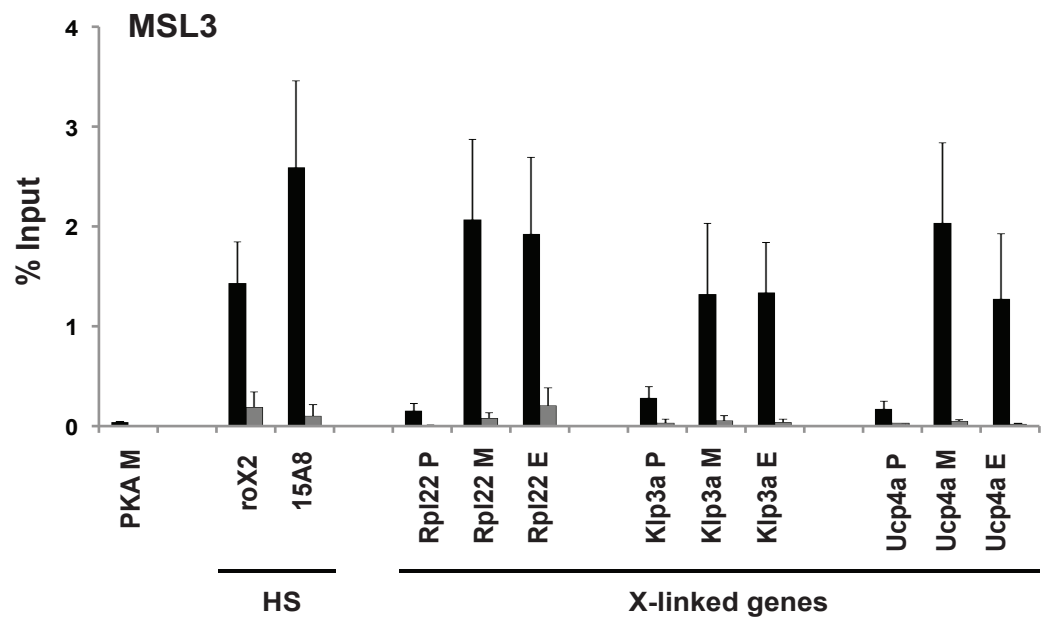


Supplementary Figure 6

A



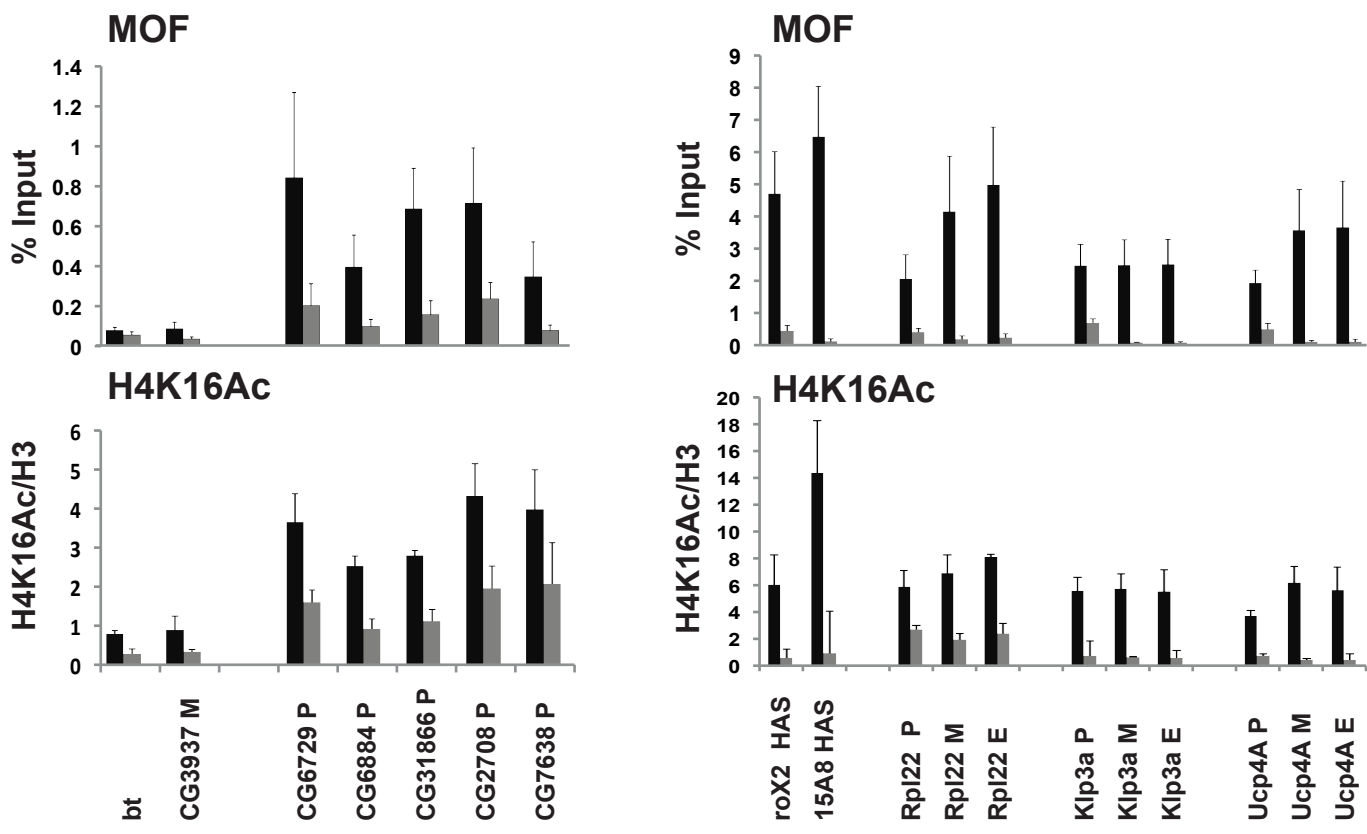
B



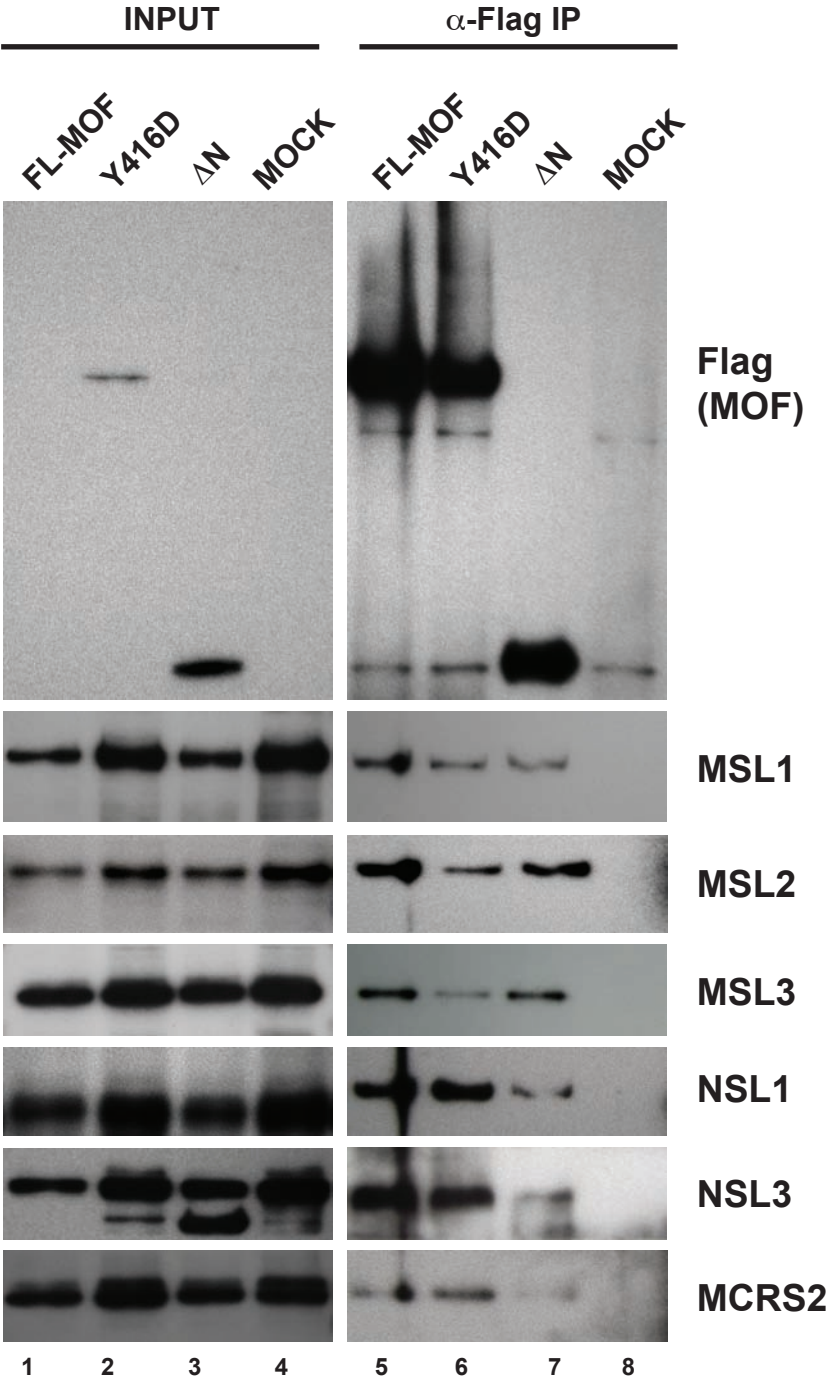
Autosomes

X chromosome

C



Supplementary Figure 7



Sample	# reads	# reads mapped	% reads mapped	coverage
<b>Male samples</b>				
MOF male wt	9,945,648	8915892	89.65	1.90
input male wt	7,143,508	6298158	88.17	1.34
H4 male wt	13,564,190	11892074	87.67	2.54
H4K16 male wt	12,567,533	11038368	87.83	2.36
H4 male <i>mof2</i>	18,391,608	15774286	85.77	3.37
H4K16 male <i>mof2</i>	18,033,157	14565538	80.77	3.11
input male <i>mof2</i>	9,934,408	8641131	86.98	1.84
<b>Female samples</b>				
MOF female wt	9,935,297	7008690	70.54	1.50
Input female wt	15,784,565	13590676	86.10	2.90
H4 female wt	12,727,549	11094654	87.17	2.37
H4K16 female wt	10,476,721	8987154	85.78	1.92

**Table S1: ChIP-seq samples analysed in the study.** This table presents the total numbers of reads sequenced from each samples, the numbers and percentages of the reads aligned by Bowtie and selected as they were mapped on one of the *D. melanogaster* chromosome and did not contain any Ns, and the genome coverage obtained from these selected mapped reads.

	All genes		Expressed genes	
	X (%)	Auto. (%)	X exp (%)	Auto exp (%)
<b>Male wt</b>				
<b>MOF</b>				
Total bound	1,285 (57.65)	4,830 (41.19)	704 (92.63)	3,132 (81.48)
gene body	898 (40.29)	50 (0.43)	569 (74.87)	27 (0.7)
promoter	387 (17.36)	4,780 (40.77)	135 (17.76)	3,105 (80.78)
no MOF	944 (42.35)	6,895 (58.81)	56 (7.37)	712 (18.52)
Total	2,229	11,725	760	3,844
<b>H4K16ac</b>				
Total bound	1,989 (89.23)	4,690 (40)	748 (98.42)	3,111 (80.93)
gene body	1,704 (76.45)	321 (2.74)	712 (93.68)	197 (5.12)
promoter	285 (12.79)	4,369 (37.26)	36 (4.74)	2,914 (75.81)
no H4K16ac	240 (10.77)	7,035 (60)	12 (1.58)	733 (19.07)
Total	2,229	11,725	760	3,844
<b>Female wt</b>				
<b>MOF</b>				
Total bound	903 (40.51)	3,862 (32.94)	620 (74.79)	2,745 (65.02)
gene body	1 (0.04)	8 (0.07)	1 (0.12)	3 (0.07)
promoter	902 (40.47)	3,854 (32.87)	619 (74.67)	2,742 (64.95)
no MOF	1,326 (59.49)	7,863 (67.06)	209 (25.21)	1,477 (34.98)
Total	2,229	11,725	829	4,222
<b>H4K16ac</b>				
Total bound	1,064 (47.73)	5,023 (42.84)	695 (83.84)	3,460 (81.95)
gene body	158 (7.09)	682 (5.82)	89 (10.74)	445 (10.54)
promoter	906 (40.65)	4,341 (37.02)	606 (73.1)	3,015 (71.41)
no H4K16ac	1,165 (52.27)	6,702 (57.16)	134 (16.16)	762 (18.05)
Total	2,229	11,725	829	4,222

**Table S2. Numbers and percentages of MOF-bound and H4K16-acetylated genes.**

Numbers and percentages of MOF-bound and H4K16-acetylated genes on the X chromosome and the autosomes in wt males (top) and females (bottom). The numbers and percentages of expressed genes are indicated in the right end columns (X exp and Auto exp).